Initial Transient Phase of Steady state Simulation: Methods of Its Length Detection and Their Evaluation in Akaroa2

A Thesis
Submitted in Partial Fulfilment
of the Requirement for the Degree
of
Master of Science in Computer Science in the
University of Canterbury
by

Behshid Ghorbani

University of Canterbury
2004

Examining Committee:

Prof. Dr Krzysztof Pawlikowski, University of Canterbury (supervisor)
Associate Prof. Dr Don McNickle, University of Canterbury (co-supervisor)
Prof. Dr Oliver Rose, University of Dresden, Germany (external examiner)

Acknowledgements

I would like to thank my supervisors Krzysztof Pawlikowski and Don McNickle for their comments, suggestions, and tireless help, and Greg Ewing with his help with the Akaroa2 package. I would like to also thank my family for their support and my friends Dan, Andreas, Jason, Josh, Nadine, Sung, Fung, and Malcolm who as a collective group ensured that I had support at any hour of the day or night. Lastly, I would like to thank Tim Bell for his patience and assistance through the duration of my studies.

Contents

Chapter 1

Introduction

The start-up or initial transient problem arises in steady state, discrete-event simulation, where a selection of non-typical initial conditions introduces bias in simulated output sequences. One way of dealing with initialisation bias is to delete a portion of the output from the beginning of the run, to eliminate the effects of bias caused by these initial conditions. To make sure that the remaining observations represent steady state behaviour, it is safer to remove more than enough observations from the beginning of the run but only to an extent that not too many good observations are removed. This research focuses on the initial transient period and the proposed statistical tests for detecting its length. Our aim here is to find a statistical test that in addition to overestimating the length of the initial transient period it detects a length as accurately as possible (in finding a length as close to the one estimated by the theory).

The background is outlined in Section 1.1 of this chapter, where two types of simulations, the initial transient period, and the ways to deal with the initial bias are described. A brief introduction to the Akaroa2 package is also presented in this section. Section 1.2 reveals the objectives of this research. The structure of this thesis is presented in Section 1.3.

1.1. Background

1.1.1. Simulation

"Simulation is the process of designing a model of a real system and conducting experiments with this model for the purpose either of understanding the behaviour of the system or of evaluating various strategies for the operation of the system" [Shannon, 1975]. Simulation, like most analysis methods, involves systems and models of them. A bank, a computer network, or a telecommunication network, are all examples of systems to be modelled.

1.1.1.1. Simulation Types

Most simulations can be classified as either *terminating* or *steady state*:

1. "Terminating simulation: a terminating simulation is one in which the model dictates specific starting and stopping conditions as a natural reflection of how the target system actually operates. As the name suggests, the simulation will terminate according to some model-specified rule or condition. For example, a store opens at 9 AM with no customer present, closes its doors at 9 PM, and then continues operation until all customers have left the store" [Kelton, Sadowski, and Sadowski, 2002].

2. Steady state simulation: "A steady state simulation, on the other hand, is one in which the quantities to be estimated are defined in the long run; i.e., over a theoretically infinite time frame" [Kelton, Sadowski, and Sadowski, 2002]. "In a steady state simulation, both the initial conditions and the length of the simulation are determined by the modeller, and the measure of interest is defined in terms of a limiting value reached, as the length of the simulation run goes to infinity. Suppose we are interested in the mean waiting time in a queueing system Theoretically, as the length of a simulation of this system approaches infinity, the distribution of waiting times become unchanged and the mean waiting time converges to a limiting value. Practically, run lengths are finite and simulation provides only a sample set of values from this distribution" [MacDougall, 1987]. An emergency room of a hospital that never really stops or restarts can be an example of a steady state simulation.

1.1.2. Initial Transient Period

In the case of the terminating simulation, because classical methods of statistical analysis can be directly applied, the analysis of the output is fairly simple and straightforward. The start-up or warm-up problem arises in steady state simulation that is conducted for performance evaluation of systems when they reach steady state. Because of that, one needs to know when to start collecting output data that represent steady state behaviour. This means that we have to know the length of the initial transient period during which data should not be collected. The focus of this research is to investigate different statistical tests proposed for detecting the length of initial transient period in steady state simulation, when the output data analysis is conducted automatically and on-line, during simulation.

Given a stationary simulation process, for which the mean, $E(X)$, and the variance, s^2, are usually unknown, the sample mean, $\bar{x}$, and the sample variance, $\hat{s}^2$ are used as their point estimators. "While there are many methods for estimating the variance parameter for such processes (see, for example, Pawlikowski [1990], Alexopoulos, and Seila [1996], Law, and Kelton [1991], Law [1983], and Wilson [1984]), they usually assume that the process has reached steady state before data collection begins. If this is not the case, then transient behaviour can have a significant impact on the estimate of $E(X)$ and s^2" [Ockerman, Goldsman, 1997]. Various methods have been proposed to deal with this problem (see, for example, Pawlikowski [1990], Ockerman, and Goldsman [1997], Goldsman, Schruben, and Swain [1994], Wilson, and Pritsker [1978], and Chance [1993]).

A common way of dealing with the initialisation bias is to collect a large number of observations in an attempt to overwhelm the initialisation effects. This method is called the extra long replication method and is described in Law and Kelton [2000]. Another method is to run the simulation for a period of time, then delete (truncate) a portion of the output from the beginning of the run to eliminate or alleviate the effects of bias. However, if the output is truncated too early, then significant bias might still be present. If it is truncated too late, then good observations are lost [Goldsman, Schruben, and Swain, 1994].

A variety of methods for analysing the output of a simulation model of a steady state simulation have been proposed (see for example Pawlikowski [1990]). "Most of these methods attempt to eliminate the initial bias by discarding the observations during the initial transient period and continuing the simulation with the system in a more representative state. However, the regenerative method (see Pawlikowski [1990]) avoids the problem of initial bias by using regeneration points. A regeneration point is a system state in which the future behaviour of the system is independent of the system's past history. It is not possible to identify regeneration points for every system, but when a system does have a regeneration point, it can be exploited to yield point and interval estimates of the system's properties. An example of a system with regeneration points can be a queueing system when there are no customers in the queue or in the service" [Hoover and, Perry, 1990].

1.1.2.1. Initial Transient Tests

In quantitative stochastic simulation, techniques that are more powerful than heuristics should be applied. After removing the transient portion of the data, an initialisation bias test should be performed to decide if the remaining data is represented of the steady state behaviour (see, for example, Pawlikowski [1990], Ockerman, and Goldsman [1997], Goldsman, Schruben, and Swain [1994], Schruben [1982], Schruben, Singh, and Tierney [1983], Yücesan [1993], and Vassilacopoulos [1989]). These initialisation bias tests are typically hypothesis tests with the null hypothesis, H_0: no initialisation bias present, and the alternative hypothesis, H_1: initialisation bias is present. In Stacey [1993], two of these tests, namely the Schruben test [Schruben, 1982] and the Yücesan test [Yücesan, 1993] were implemented and their performance investigated. The result of Stacey's investigation showed that the Yücesan test detected a length closer to the theory than the one detected by the Schruben test (see Stacey [1993]). In Goldsman, Schruben, and Swain [1994], the authors proposed a family of tests that extended the tests proposed by Schruben [1982] and Schruben, Singh and Tierney [1983]. We look at these tests in more detail in Chapter 2.

1.1.3. Single and Multiple Replications in Parallel

To obtain final results with small errors, very long simulation runs may be needed. This may lead to a very long simulation time. In this situation, we can speed up the simulation by using distributed simulation. We can distinguish two approaches to parallel execution of simulated processes on multi-processor computers or multiple computers of local networks:

1. SRIP (Single Replication In Parallel), which is shortening the execution time of a simulation by reducing the complexity of the simulation model. This is done by partitioning the model into sub models, making the simulation of sub-models on different processors simpler and faster. The main problem with this scenario is that not all the systems can be partitioned into truly independent subsystems.

2. MRIP (Multiple Replications In Parallel). A simulation can be sped up if observations are produced in parallel, by multiple processors running statistically independent replications of the same simulation. Such processors can be viewed as simulation engines working in a team and producing one common sample of output data. Observations generated by different simulation engines, but representing values of the same performance measure, are submitted to a global analyser that is responsible for their statistical analysis. The current statistical error of results should be analysed at consecutive checkpoints. The analysis of each performance measure is continued as long as the statistical error of its estimate does not drop below an assumed acceptable level. All simulation

engines should operate until the analysis of all performance measures are finished. At that instance of time all simulation engines are stopped. The global analyser the produce the final results.

An example of implementation of MRIP is the Akaroa2 package, which has been designed within the Akaroa project at the university of Canterbury since 1991. A sequential version of the Schruben test has been implemented in Akaroa2 to remove the observations in the initial transient period. We also used this package to implement a sequential version of the statistical tests that are listed in Chapter 3 to detect the length of the initial transient period.

1.2. Research Objective

Initialisation bias can be a major source of errors when estimating the steady state values of system performance measures. Despite research on the initial transient period and numbers of tests proposed to detect it, no satisfactory conclusions on the most efficient or accurate tests have been attained. The objective of this research is:

1. To find a test that detects a length of the initial transient period close to the one suggested by theoretical results, such as the relaxation time or the algorithm by Kelton and Law [1983].

2. To find a test that detects a length of initial transient period, such that the output data process after this period can be considered as being in steady state.

We consider three different queueing models, for the evaluation of the performance of these tests: M/M/1/8, M/Erlang$_4$/1/8, and M/Pareto$_{a=2.1}$/1/8 with the coefficients of variation of the service times equal to 1, 0.5, and 2.18, respectively. The performance of the best tests has been also evaluated on six artificially generated stochastic processes.

1.3. Research Structure

Chapter 2 of this research provides a comprehensive survey of the approximations based on the theory and the statistical tests proposed to detect the length of the initial transient period. Chapter 3 presents the results of the experiments conducted to investigate the issues listed in Section 1.2: to find the most efficient and accurate test for detecting the length of the initial transient period. The conclusions are contained in Chapter 4.

Chapter 2

Theoretical and Statistical Studies of Methods for the Detection of the Initial Transient

In steady state simulation, the output data collected at the beginning of the simulation during the initial warm-up period may not represent steady state behaviour. If they are included in further analysis, they can cause a significant bias of the final results. A common way of dealing with initialisation bias is to run the simulation for a period of time, then delete a portion of the output from the beginning of the run to eliminate the effects of the initialization bias. Therefore, we need to know when to start collecting output data that are representative of steady state behaviour. This means that we have to know the length of the initial transient period during which data should not be collected. However, there can be some problems with this. If the output is truncated too early, then significant bias may still be present. Also if it is truncated too late, then good observations are lost. Particularly in simulations based on the multiple replications in parallel (MRIP) scenario, if we delete too many observations from each replication, then the speed up will be reduced. So how do we know how many observations to delete? There are many statistical tests that have been proposed to find the length of the initial transient period. Selecting the most efficient and accurate ones is the main part of our research.

In this chapter we first look into some theoretical studies that provide estimates of the initial transient period, followed by investigation of some of the statistical tests that have been proposed to find the length of this period. Our aim here is to find a statistical test that in addition to overestimating the length of the initial transient period it detects a length as accurate as possible (in the sense of finding a length as close to the one obtained by the theory). Of course, such comparisons can only be made in queueing systems for which the theoretical results are available.

2.1. Theoretical Measures of the Length of the Initial Transient Period

In this section, we look at two well-known theoretical approximations of the length of the initial transient period. One of them is the expected waiting time of the *n*th customer. This can be exactly calculated for the M/M/s/8 queueing system. In our research, we will only look at the expected waiting time of the *n*th customer for the M/M/1/8 queues. The other theoretical measure is the relaxation time, which is the rate at which the mean queue lengths or the mean delays tend to their steady state[1]. Both these measures are more precisely defined in Sections 2.1.1 and 2.1.2.

2.1.1. The Actual Waiting Time of the *n*th Customer for M/M/1/8 Queueing Systems

Kelton and Law [1983] studied the transient behaviour of M/M/s queues with different number of customers presented at time zero and carried out an analysis in discrete time (i.e., indexing by customer number), rather than in the continuous-time framework. Using these results, they examined the effect of the initial condition on the nature of the convergence of the exact expected waiting time in the queue by the *n*th arriving customer to its steady state value.

[1] The relaxation time can be introduced as the measure of convergence for any moment but it has been only used for the measure of means in our research.

In our research, we investigate only the M/M/1/8 queueing system[2]. Let X_n be the number of customers present in the system at time T_n where the nth "new" customer arrives. Let us define $P_k(n,i) = P(X_n = i \mid k$ customers already present at time 0), a = ?/(?+1), b = 1 - a, and r the system load. The expected waiting time of the nth arriving customer in M/M/1/8 queue that is initially empty and idle can be calculated using the following algorithm by Morisaku [1976].

1. (For n = 1) Set $P_0(1,1) = 1$.
2. (For n >= 2)
    a. Set $P_0(n,n) = a^{n-1}$.
    b. In the order $i = n - 1$, $i = n-2$, …, $i = 2$, set $P_0(n, i) = a\, P_0(n-1, i-1) + b\, P_0(n, i+1)$. (Omit this step if n = 2).
    c. Set $P_0(n, 1) = \frac{1}{r} P_0(n, 2)$.

From this the expected waiting time in the M/M/1/8 queue of the nth customer can be easily calculated as

$$\mathrm{E}(W_q^{(n)}) = \frac{1}{m} \sum_{i=2}^{k+n} (i-1) \mathrm{P}_k(n,i),$$

The steady state expected waiting time in the queue for a customer, w, can also be calculated by

$$w = \frac{l}{(m-l)^2} \mathrm{P}_0 .$$

[2] Because of similar transient behaviour of M/M/s queueing systems, see the results by Kelton and Law [1983], our investigation was limited to M/M/1 queueing systems.

Here, P_0 is the steady state probability that zero customers are presented in the system (in queue or service) and r, m and l are the system load, service rate and arrival rate, respectively. The value for P_0 is 1-r (see Gross and Harris [1974]).

Figure 1 shows how the expected waiting time of the nth customer tends toward its steady state value w for an M/M/1/8 queueing system with system load 0.9 starting from no customers at time zero.

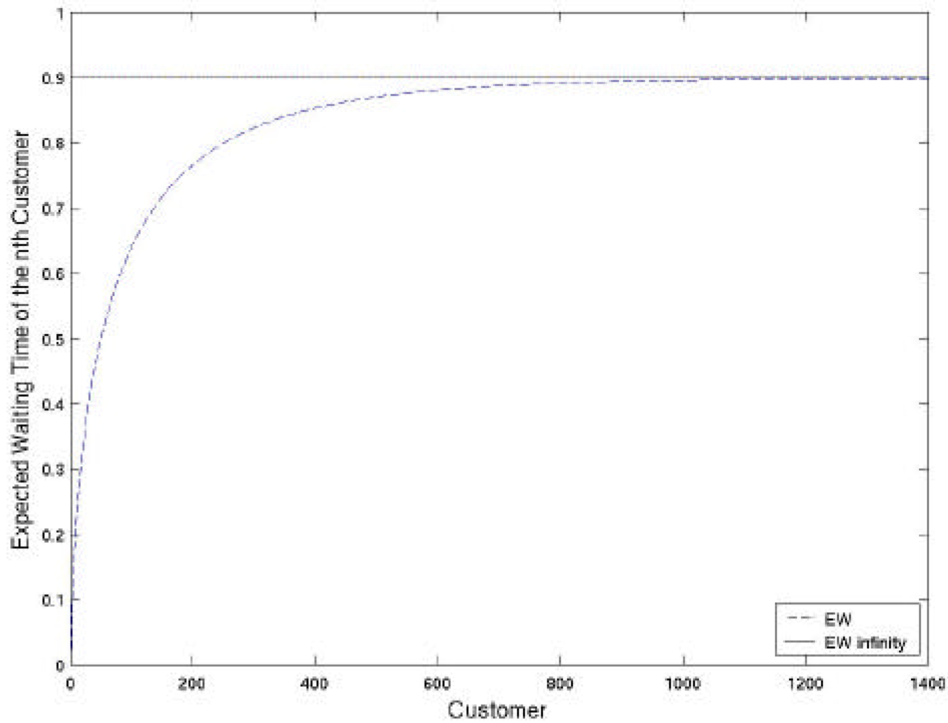

Figure 1. Expected Waiting Time of the *n*th Customer in an M/M/1/8 Queueing System with System Load 0.9.

The smallest point beyond which all values of the expected waiting times of the nth customer fall within ($p\times100$)% of w, can be calculated as

$$n(p) = \min\left\{n_0 \geq 1 : \left|\mathrm{E}(W_q^{(n)}) - w\right| \leq pw \ \textit{for all}\ n \geq n_0\right\},$$

where w is the steady state expected waiting time in the queue and p is the permissible relative residue between 0 and 1.

For example given the commonly used value for p of 0.02 (i.e. when the mean queue lengths are within 2% of their steady state values), the length of the initial transient period with system load 0.9 is $n(0.02) = 615$.

The table and graph that follow show the number of observations in the initial transient period using the Kelton and Law algorithm, for p = 0.005, 0.01, 0.02 and 0.05 for different system loads.

By looking at Figure 2, as could be expected, with smaller values of p (for a more precise results) the initial transient period is longer and so more observations would need to be deleted, if this is a requirement in steady state analysis.

Load \ *p*	0.005	0.01	0.02	0.05
0.05	4	4	3	3
0.10	5	4	4	3
0.15	6	5	5	4
0.20	8	7	6	5
0.25	9	8	7	5
0.30	11	10	8	6
0.35	14	12	10	8
0.40	17	15	12	9
0.45	22	18	15	11
0.50	27	23	19	14
0.55	35	30	24	17
0.60	47	39	32	22
0.65	64	53	43	30
0.70	91	75	60	42
0.75	137	113	90	61
0.80	222	182	144	98
0.85	411	336	265	178
0.90	961	784	615	410
0.95	3989	3244	2538	1677

Table 1. The Length of Initial Transient Period Measured in Terms of the Number of Observations for the M/M/1/8 Queue using Kelton and Law Algorithm.

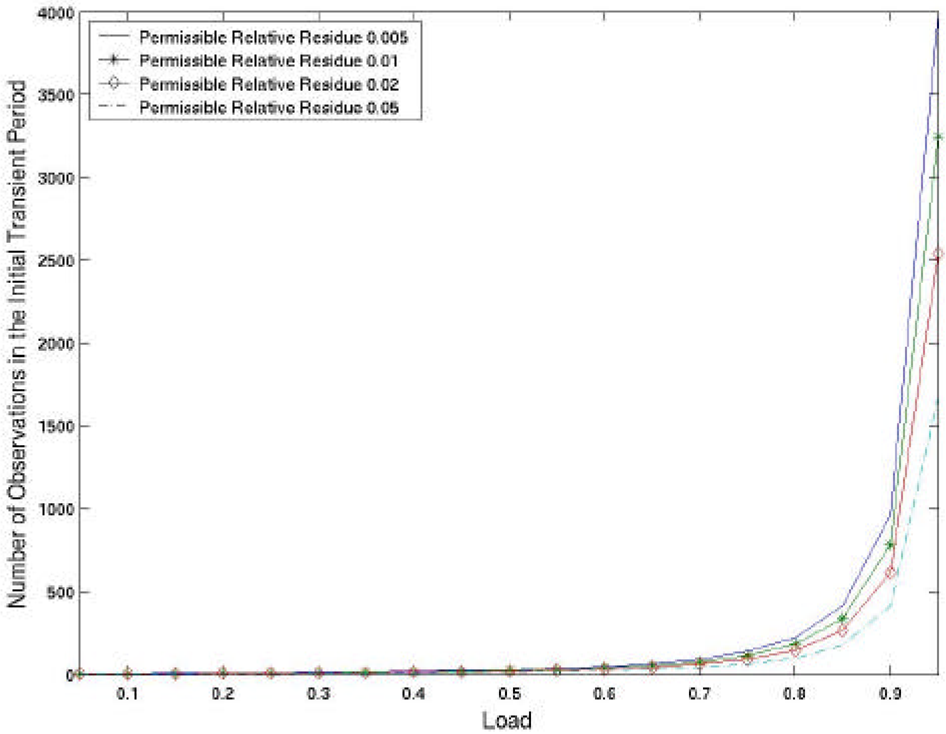

Figure 2. Length of the Initial Transient Period with Different Permissible Relative Residues in an M/M/1/8 Queue.

2.1.2. Relaxation Time

Morse [1955] first introduced the idea of relaxation time for the M/M/1/8 system. His derivation was based on an approximating expression for the correlation function of the queue length process X_t, where X_t is the queue length at time t, using the transient behaviour of the state probabilities for various initial conditions. He indicated that the transient behaviour the mean queue length, E[X_t], can be measured by the relaxation time τ, defined as the mean time required for E[X_t] to differ from its steady state mean value L by less than (1/e)100%. He also showed that this relaxation time increases at the rate of $(1\text{-}r)^{-2}$ as the utilization factor approaches 1, whereas L increases by $(1\text{-}r)^{-1}$, so the relaxation time of the mean queue length is more sensitive to saturation than is the mean queue length itself. Morse approximated the relaxation time of the mean queue length by

$$t_{Morse} = \frac{2l}{(m-l)^2} = \frac{2r}{m(1-r)^2},$$

where r, m and l are the system load, service rate and arrival rate, respectively.

Karlin and McGregor [1957] proved that for many queueing systems, the rate of convergence of the mean queue length to its steady state characteristics, eventually becomes dominated by an exponential term of the form $\exp(-t/t)$ where t is the relaxation time.

In 1969, Cohen [1969] introduced a different approach to derive the expression for relaxation time of the mean queue length in GI/G/1/8 queueing systems by considering the behaviour of $P_{ij}(t) = P(X_{y+t}=j \mid X_y=i)$, the transition probability from state i to state j, with the asymptotic relations for $E(X_t \mid X_0 = k)$ as $t \to \infty$. Cohen's relaxation time was

$$t_{Cohen} = \frac{1}{m(1-\sqrt{r})^2}.$$

As formulated by Pawlikowski [1990], the initial transient period can be considered as being over after the time t_p = -t ln p, where p is the permissible relative residue of the initial state, 0 < p < 1. Thus, assuming p = 0.02, we find that at $t = 4t$ the mean queue length is within 2% of its steady state value. In other words, the estimates obtained on the basis of output data collected from that point in time should be biased by the initial state of simulation by less than 2% [Pawlikowski, 1990].

Using Cohen's expression, the time required by the expected number of customers in the system to reach its steady state value within less than 2% equals $4t_{Cohen} = 4/(m(1-\sqrt{r})^2)$. Odoni and Roth [1981] argued that as Cohen's expression time constant is based on asymptotic ($t \to \infty$) arguments, the result of this calculation is considerably larger than the amount of time it actually takes the expected number of customers to be within 2% of its steady state value. They also discussed that the time constant proposed by Morse differs by a factor of $2r/(1+\sqrt{r})^2$ from Cohen's expression and could take values as much as 50% greater than t_{Cohen} as $r \to 1$ [Odoni and Roth, 1981]. These results will be discussed later.

Using the diffusion approximation for GI/G/1/8 queues under heavy traffic, Newell [1971] proposed the following formula for the relaxation time of the mean queue length :

$$t_{Newell} = \frac{1}{m(1-r)^2},$$

which was generalized to the formula below by Odoni and Roth [1981].

$$t_{Newell} = \frac{(r\, C_A^2 + C_S^2)}{m(1-r)^2},$$

where C_A and C_S are the coefficients of variation for the inter-arrival and service times, respectively.

Mori [1974] developed a numerical technique for estimating the transient behaviour of the expected waiting time for GI/G/1/8 systems based on a recursive relationship involving waiting times of successive customers. He compared this estimation for the general queueing systems (GI/G/1/8) with the transient behaviour of the M/M/1/8 and M/D/1/8 queueing systems (as the exact values of the mean waiting time for both these queueing systems can be easily calculated). His expression for calculating the relaxation time of the expected waiting time was

$$t_{Mori} = \frac{(C_A^2 + r\,C_S^2)}{(1-r)^2}\,.$$

Odoni and Roth [1981] also obtained an approximation to the relaxation time of the mean queue length within Markovian queueing systems[3] which was of the form

$$t_{Odoni\&\,Roth} = \frac{(1+\sqrt{r}\,)^2(C_A^2 + C_S^2)}{2.8m(1-r)^2} = \frac{(C_A^2 + C_S^2)}{2.8m(1-\sqrt{r}\,)^2}\,.$$

They reported that this was consistent in form with Cohen’s approximation for the relaxation time for M/M/1/8 queueing systems [Odoni and Roth, 1981]. They found that the relaxation time varies directly with square powers of C_A and C_S. In addition, as the system approaches saturation, the system requires a greater amount of time to approach equilibrium; thus t would be expected to vary directly with some power of 1/(1-?). Using the expression proposed by Odoni and Roth [1981], the relaxation time can be calculated for the M/M/1/8, M/Erlang$_k$/1/8 and M/Pareto$_{\alpha=2.1}$/1/8 queueing systems where $C_A^2 = 1$ and C_S^2 is equal to 1, 1/k and

[3] “The Markovian class includes systems in which inter-arrival and service times can be represented as exponential, Erlangian, and hyperexponantial” Odoni and Roth [1981].

1/(a(a-2)) for the M/M/1/8, M/Erlang$_k$/1/8 and M/Pareto$_{\alpha=2.1}$/1/8 queueing systems, respectively.

Jackway and deSilva [1992] used an approximate translation to convert a relaxation time estimate to one based on the number of observed service completions. Thus, by dividing the length of initial transient of the mean queue length, equal 4*t* (for *p* = 0.02), by 1/*m*, the length of the initial transient period of the mean number of service completions can be calculated [Stacey, 1993].

All these results can be used to calculate the length of the initial transient period but their applications are restricted to a relatively narrow family of queueing systems. Therefore, they can be used only as a (theoretical) reference for studying performance of other methods of transient analysis that can be applied to any model.

The following tables show the comparison of the number of observations in the initial transient period of the mean queue length using the above mentioned relaxation time estimates and the algorithm proposed by Kelton and Law [1994], considering an M/M/1/8 queue with the permissible relative residue of 0.5%, 1%, 2%, and 5%. As the following graphs (Figures 3-6) show, using the relaxation time expression proposed by Cohen gives the highest number of observations in the initial transient period, whereas the algorithm given by Kelton and Law results in the smallest number of observations in the initial transient period. It can also be observed that applications of different relaxation time formulas for calculating the number of observations in the initial transient period of the mean queue length produce results of the same order, for any of the three considered values of p: p=0.005, 0.01 and 0.02.

Load \ p=0.005	Kelton & Law	Morse	Cohen	Newell	Odoni & Roth
0.05	4.0	0.6	8.8	6.2	6.3
0.10	5.0	1.3	11.3	7.2	8.1
0.15	6.0	2.2	14.1	8.4	10.1
0.20	8.0	3.3	17.3	9.9	12.4
0.25	9.0	4.7	21.2	11.8	15.1
0.30	11.0	6.5	25.9	14.1	18.5
0.35	14.0	8.8	31.8	16.9	22.7
0.40	17.0	11.8	39.2	20.6	28.0
0.45	22.0	15.8	48.9	25.4	34.9
0.50	27.0	21.2	61.8	31.8	44.1
0.55	35.0	28.8	79.4	40.6	56.7
0.60	47.0	39.8	104.3	53.0	74.5
0.65	64.0	56.2	141.2	71.4	100.8
0.70	91.0	82.4	198.7	100.1	141.9
0.75	137.0	127.2	295.3	148.4	210.9
0.80	222.0	212.0	475.5	238.5	339.7
0.85	411.0	400.4	870.1	435.8	621.5
0.90	961.0	954.0	2012.6	1007.0	1437.6
0.95	3989.0	4028.0	8266.6	4134.0	5904.7

Table 2. Length of the Initial Transient Period Measured by the Relaxation Time and Kelton and Law Algorithm for *p*=0.005 in M/M/1/8 Queueing System.

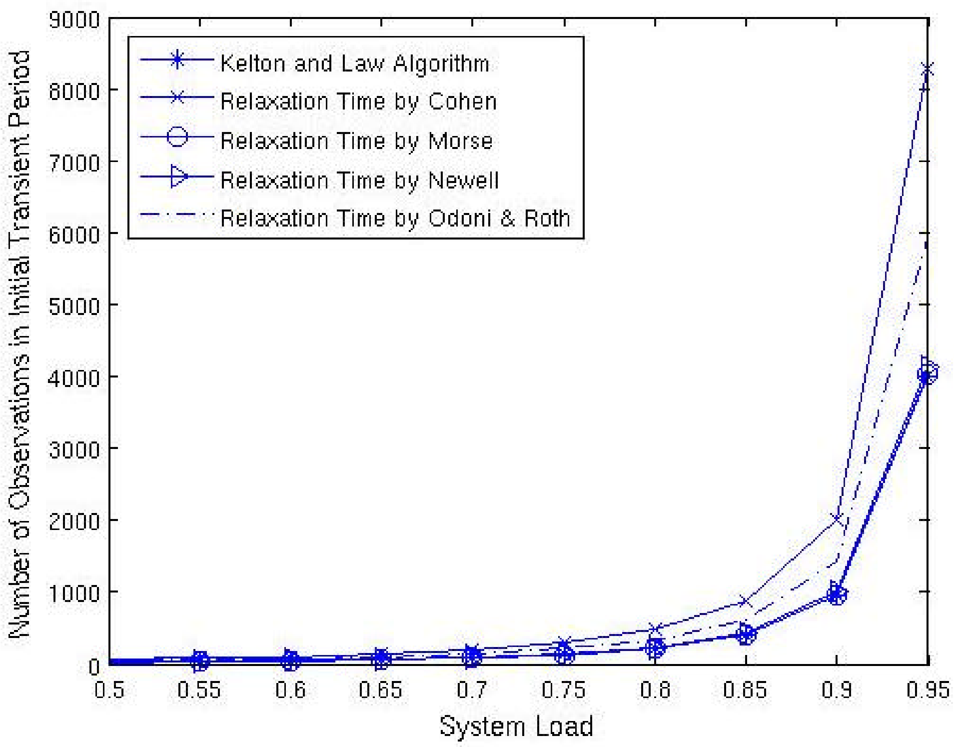

Figure 3. Length of the Initial Transient Period Measured by the Relaxation Time and Kelton and Law Algorithm for *p*=0.005 in M/M/1/8 Queueing System.

p=0.01 / Load	Kelton & Law	Morse	Cohen	Newell	Odoni & Roth
0.05	4.0	0.5	7.6	5.4	5.5
0.10	4.0	1.1	9.8	6.2	7.0
0.15	5.0	1.9	12.3	7.3	8.8
0.20	7.0	2.9	15.1	8.6	10.8
0.25	8.0	4.1	18.4	10.2	13.1
0.30	10.0	5.6	22.5	12.2	16.1
0.35	12.0	7.6	27.6	14.7	19.7
0.40	15.0	10.2	34.1	17.9	24.3
0.45	18.0	13.7	42.5	22.0	30.3
0.50	23.0	18.4	53.6	27.6	38.3
0.55	30.0	25.0	68.9	35.2	49.2
0.60	39.0	34.5	90.5	46.0	64.7
0.65	53.0	48.8	122.5	62.0	87.5
0.70	75.0	71.6	172.4	86.9	123.2
0.75	113.0	110.4	256.3	128.8	183.1
0.80	182.0	184.0	412.7	207.0	294.8
0.85	336.0	347.6	755.2	378.2	539.4
0.90	784.0	828.0	1746.8	874.0	1247.7
0.95	3244.0	3496.0	7174.8	3588.0	5124.9

Table 3. Length of the Initial Transient Period Measured by the Relaxation Time and Kelton and Law Algorithm for p=0.01 in M/M/1/8 Queueing System.

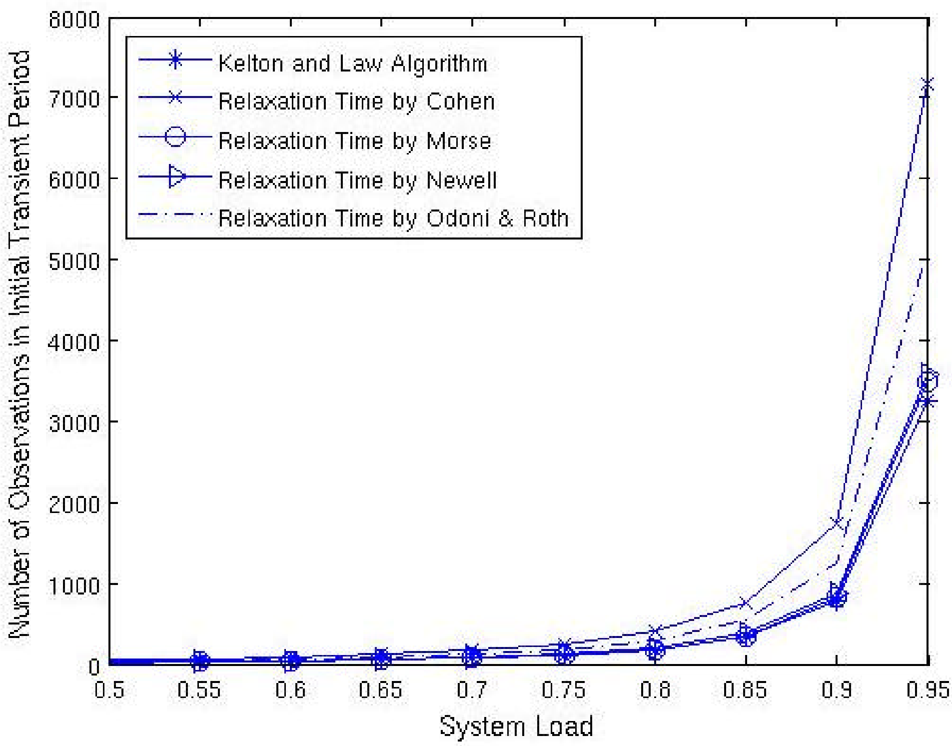

Figure 4. Length of the Initial Transient Period Measured by the Relaxation Time and Kelton and Law Algorithm for p=0.01 in M/M/1/8 Queueing System.

p=0.02 / Load	Kelton & Law	Morse	Cohen	Newell	Odoni & Roth
0.05	3.0	0.4	6.6	4.7	4.7
0.10	4.0	1.0	8.6	5.4	6.1
0.15	5.0	1.7	10.7	6.4	7.6
0.20	6.0	2.5	13.1	7.5	9.3
0.25	7.0	3.6	16.0	8.9	11.4
0.30	8.0	4.9	19.6	10.6	13.9
0.35	10.0	6.6	24.0	12.8	17.1
0.40	12.0	8.9	29.6	15.6	21.1
0.45	15.0	11.9	36.9	19.2	26.3
0.50	19.0	16.0	46.6	24.0	33.3
0.55	24.0	21.7	59.9	30.6	42.7
0.60	32.0	30.0	78.7	40.0	56.2
0.65	43.0	42.4	106.5	53.9	76.0
0.70	60.0	62.2	149.9	75.6	107.0
0.75	90.0	96.0	222.9	112.0	159.1
0.80	144.0	160.0	358.9	180.0	256.3
0.85	265.0	302.2	656.7	328.9	469.0
0.90	615.0	720.0	1518.9	760.0	1084.9
0.95	2538.0	3040.0	6239.0	3120.0	4456.4

Table 4. Length of the Initial Transient Period Measured by the Relaxation Time and Kelton and Law Algorithm for *p*=0.02 in M/M/1/8 Queueing System.

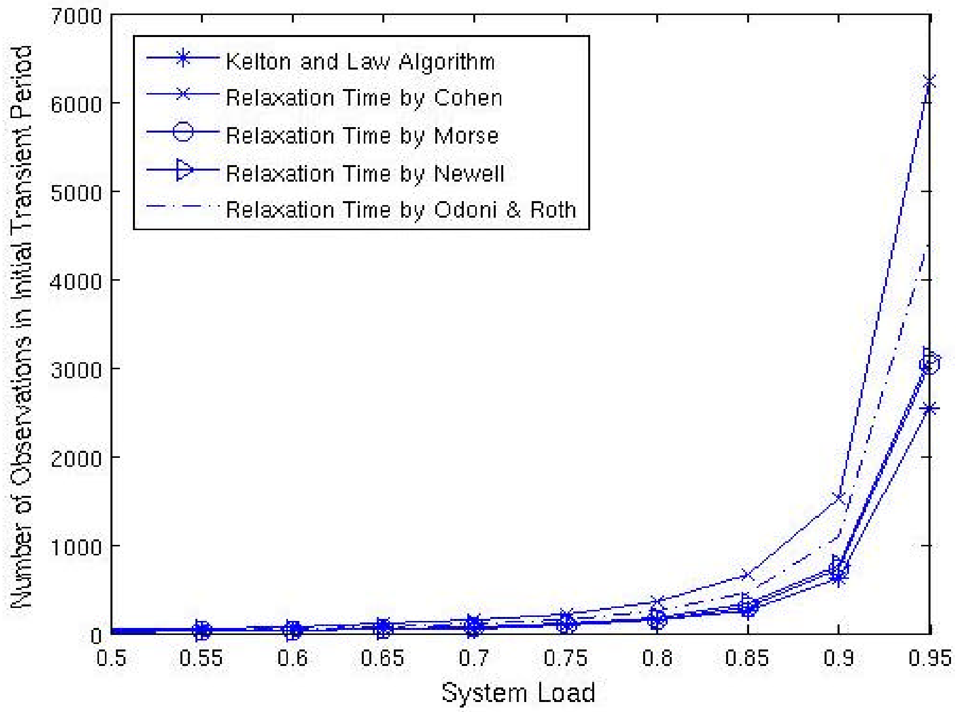

Figure 5. Length of the Initial Transient Period Measured by the Relaxation Time and Kelton and Law Algorithm for *p*=0.005 in M/M/1/8 Queueing System.

Load \ p=0.05	Kelton & Law	Morse	Cohen	Newell	Odoni & Roth
0.05	4.0	0.5	7.6	5.4	5.5
0.10	4.0	1.1	9.8	6.2	7.0
0.15	5.0	1.9	12.3	7.3	8.8
0.20	7.0	2.9	15.1	8.6	10.8
0.25	8.0	4.1	18.4	10.2	13.1
0.30	10.0	5.6	22.5	12.2	16.1
0.35	12.0	7.6	27.6	14.7	19.7
0.40	15.0	10.2	34.1	17.9	24.3
0.45	18.0	13.7	42.5	22.0	30.3
0.50	23.0	18.4	53.6	27.6	38.3
0.55	30.0	25.0	68.9	35.2	49.2
0.60	39.0	34.5	90.5	46.0	64.7
0.65	53.0	48.8	122.5	62.0	87.5
0.70	75.0	71.6	172.4	86.9	123.2
0.75	113.0	110.4	256.3	128.8	183.1
0.80	182.0	184.0	412.7	207.0	294.8
0.85	336.0	347.6	755.2	378.2	539.4
0.90	784.0	828.0	1746.8	874.0	1247.7
0.95	3244.0	3496.0	7174.8	3588.0	5124.9

Table 5. Length of the Initial Transient Period Measured by the Relaxation Time and Kelton and Law Algorithm for p=0.05 in M/M/1/8 Queueing System.

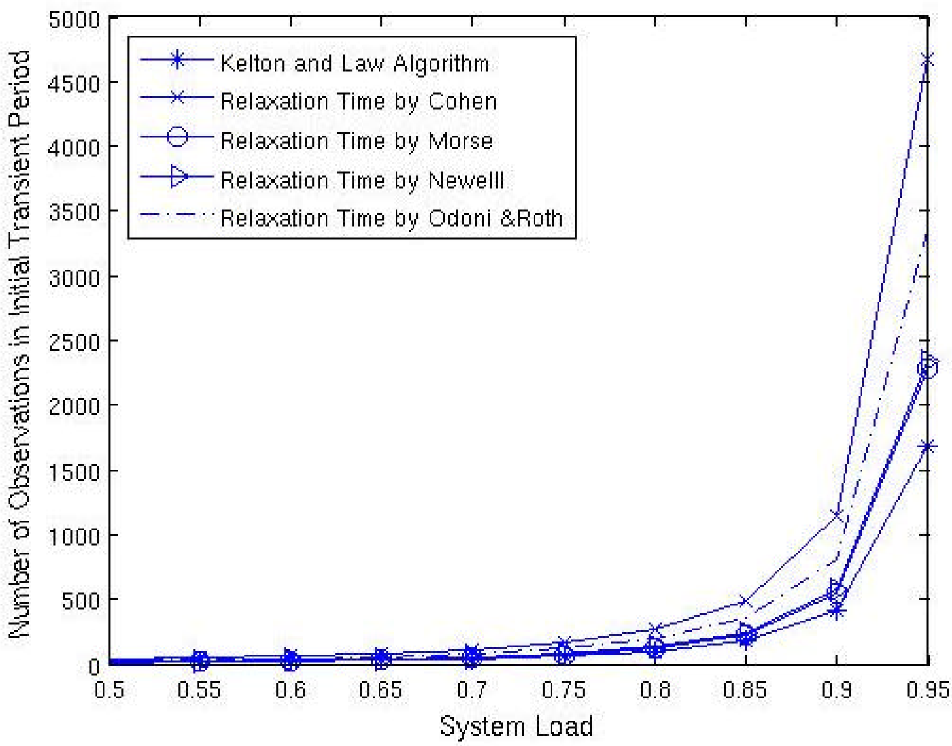

Figure 6. Length of the Initial Transient Period Measured by the Relaxation Time and Kelton and Law Algorithm for p=0.05 in M/M/1/8 Queueing System.

Figure 7 shows that by changing the permissible relative residues of 0.001 to 0.1, the length of the initial transient period calculated by the Kelton and Law algorithm can be increased almost 6 times (in heavily loaded systems) and it will even produce a length larger than the one approximated by Cohen's relaxation time.

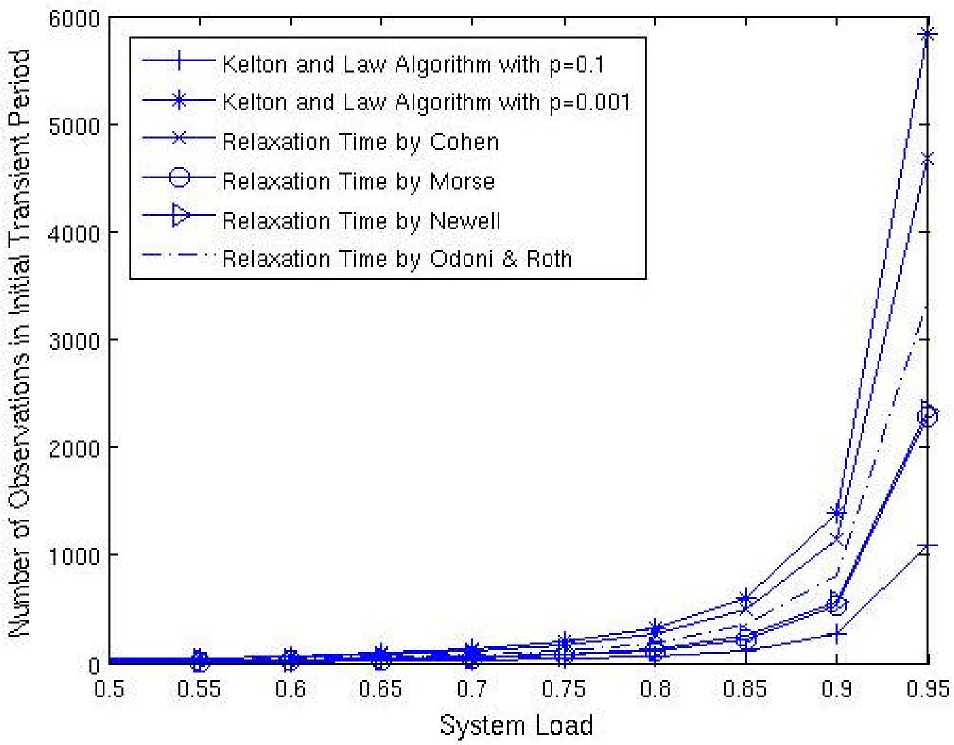

Figure 7. Length of the Initial Transient Period Measured by the relaxation (*p* = 0.05) time and Kelton and Law Algorithm (*p* = 0.001 and *p* = 0.1) in M/M/1/8 Queueing System.

2.2. Statistical Measures of the Length of the Initial Transient Period

More precise approximations of the length of the initial transient period could be obtained by using heuristics and statistical tests, some of which are studied in the following sections. Figure 8 shows a realisation of the simulated waiting time of the *n*th customer in an M/M/1/8 queue with system load 0.9. This figure shows that it is very difficult to detect the length of the initial transient period visually.

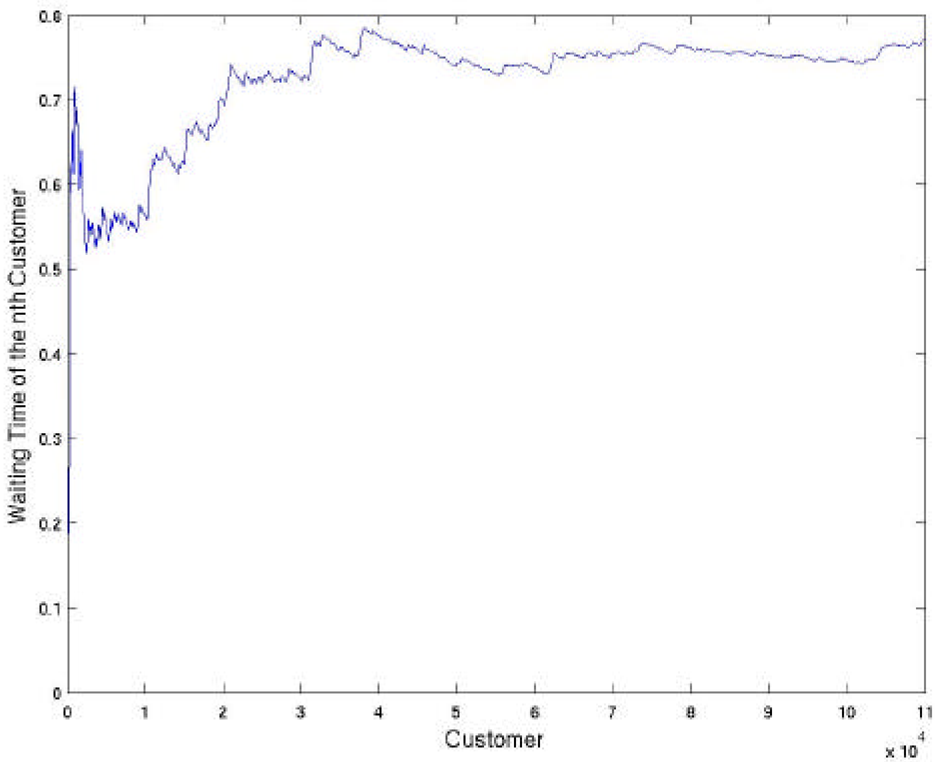

Figure 8. The Simulated Waiting Time of the *n*th Customer in an M/M/1/8 Queue with Load 0.9.

We may use a heuristic rule based on the observed simulation output data to estimate the length of the initial transient phase (for a survey of the heuristics see for example Pawlikowski [1990]). "However, our experience has shown that no heuristic can be considered relatively robust and those that we have studied have frequently produced inaccurate results" [McNickle, Stacey, and Pawlikowski, 1993]. The heuristic rules, because of their inaccuracy, can be only used as the first rough estimates of the length of initial transient. The following outlines some of the heuristics to detect the length of the initial transient period.

White [1994, 1997] and Spratt [1998] proposed the MSER (Marginal Standard Error Rule) and the MSER-5 heuristics based on selecting a truncation point that minimizes the width of the marginal confidence interval[4] about the truncated sample mean. In other words, removing initial observations that are far from the sample mean reduces bias, but only to the extent of the resulting reduction in the calculation of the confidence interval half-width. As described in Pawlikowski [1990] this accomplishment can be difficult as to reduce the half-width of confidence interval, more observations need to be deleted causing the variance to be increased. Thus, given a finite output series $X_1, X_2, X_3, ..., X_n$, the optimal truncation point for the sequence is

$$d^* = \arg\min_{n>>d\geq 0}\left[\frac{1}{(n-d)^2}\sum_{i=d+1}^{n}(X_i - \overline{X}_{n,d})^2\right]$$

where argmin returns the index of a minimal element of the list and $\overline{X}_{n,d}$ is the sample mean of the observations up to the truncation point *d*. MSER-*m* uses the series of $b = \left\lfloor \frac{n}{m} \right\rfloor$ batch averages instead of the raw output series $X_1, X_2, X_3, ..., X_n$ used in the MSER rule.

Linton and Harmonosky [2002] compared the performance of five initialisation bias detective methods namely, Welch's method [Welch, 1983], extra long replication method [Law and Kelton, 2000], relaxation time Heuristic, Kelton and Law's method [Kelton and Law, 1983] and MSER [White, 1997]. The last four methods were described in the previous sections. In the Welch method, the

[4] Marginal confidence intervals do not account for the correlation between the parameters estimated so examining only marginal confidence intervals can sometimes be misleading if there is a strong correlation between several parameter estimates.

truncation point is chosen by averaging observations across several replications and then visually choosing a truncation point based on the averaged run. For more details on Welch's method see [Welch, 1983] and [Law and Kelton, 1991]. For each method, except Extra Long Replication and the relaxation time Heuristic, the number of observations in the initial transient period and the time for the system to reach steady state were measured. In all cases except the Extra Long Replication method, each model was run for only 10 independent replications at 5000 time units. For the Extra Long Replication Method, the run time was increased to 15000 time unit. By looking at their results we can see that the length of the initial transient period detected by the MSER method was the longest, over three times longer than the method proposed by Welch and one and half times longer than the algorithm by Kelton and Law. They concluded that the Welch method and the relaxation time Heuristic were the most accurate methods. However, they chose the Welch method (despite of having to choose a length visually) as the most practical rule considering that it is not based on any assumptions about the type of system being modelled.

Robinson [2002] described a new method based on the principle of statistical process control (SPC) for estimating the length of the initial transient period. In manufacturing systems, some special circumstances may cause a process to vary according to some fixed distribution about a constant mean. Robinson expressed a close relationship between this concept of SPC and those of transience and steady state in simulation output analysis. Similar to the Welch method, the SPC method is based upon visual inspection of a time-series of the data. The disadvantage of the SPC method is that there are rigorous assumptions such as the normality and independence in the data that have to be made.

In the following sections, we will give a detailed summary of some statistical tests to detect the length of the initial transient period. The next section describes a sequential method used by some of these statistical tests.

Sequential Method of Detecting the Length of Initial Transient Period

With a given sequence of observations, statistical tests can show if this sequence of observations is in steady state or in the initial transient phase. To detect the length of the initial transient period, a sequential method was presented in Pawlikowski [1990]. The following flowchart presents this sequential procedure, which has been used in the implementation of the initial transient detector in Akaroa2. Such a sequential approach has been used in implementations of mostly all of the tests described in Chapter 3.

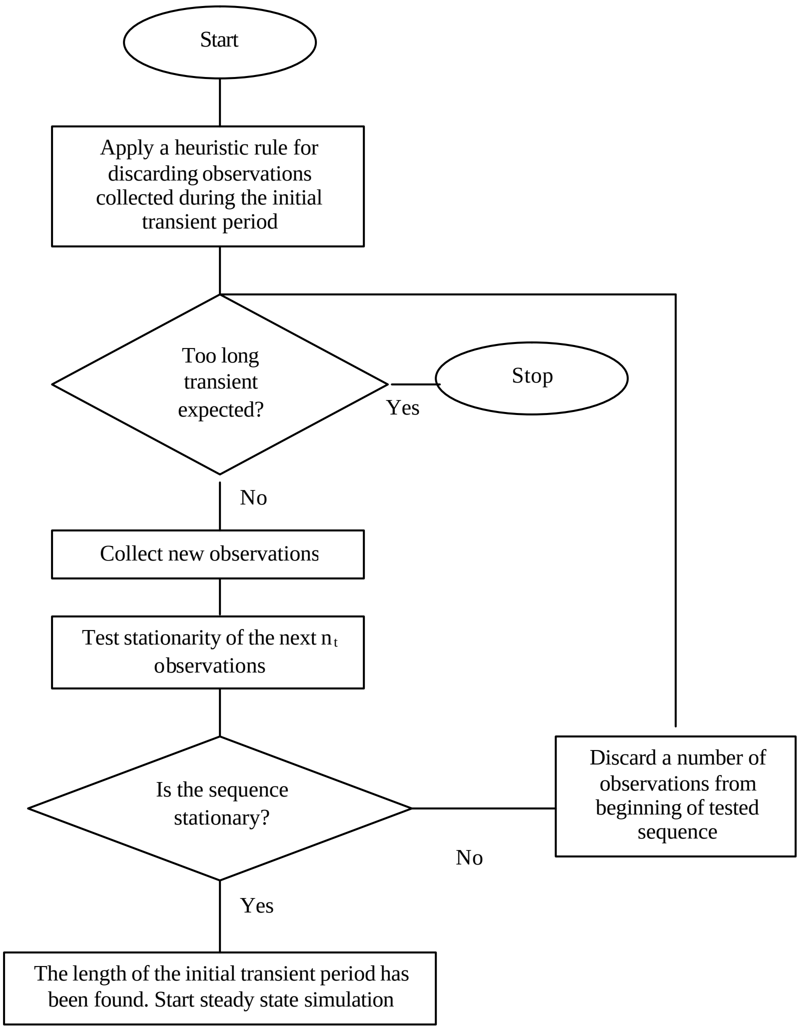

Figure 9. Sequential Method to Detect the Length of the Initial Transient Period [Pawlikowski, 1990].

2.2.1. Schruben Tests

The Schruben tests were introduced by Schruben [1981] and improved later by Schruben, Singh, and Tierney [1983]. These tests detect the initialisation bias in the mean of a simulation output series using a hypothesis-testing framework. Considering the simulation output sequence, $X_1,...,X_n$, the transient mean function $E(X_i)$ is defined as:

$$E(X_i) = E(X) \times (1 - a_i),$$

where $E(X)$ is the steady state mean and a_i represents changes in the output mean due to the initial transient bias. Therefore, there is no initial transient bias if $E(X_i) = E(X)$ for all i (i.e. if $a_i = 0$ for all i). The null hypothesis is that the output mean does not change throughout the simulation run, i.e. H_0: $a_i = 0$ for all i. The alternative hypothesis is, H_1: a_i = a specified function of i.

Schruben et al [1983] showed that these tests are asymptotically optimal based on cumulative sums of deviations about the sample mean, $S_k = \overline{X_n} - \overline{X_k}$ with $\overline{X_k} = 1/(k\sum_{i=1}^{k} X_i)$ and the convergence of a standardized time series[5] $B_t = ([nt]S_{[nt]})/\sqrt{n}; \quad 0 \le t \le 1$, to the Brownian bridge process with zero mean and variance equal to 1. A Brownian bridge process is a zero mean Guassian process on the unit interval between 0 and 1 with continuous paths. The test statistic is

$$T = \sum_{k=1}^{n_t} k(1 - \frac{k}{n_t})\left[\overline{X}_{n_t} - \overline{X}_k\right],$$

[5] "The method of standardised time series, originally proposed by Schruben [1983], relies on the convergence of standardized random processes to a Wiener random process with independent increments, also known as a Brownian bridge process. It is an application of the theory of dependent random processes and its functional central limit theorem, which is a generalization of the central limit theorem" [Pawlikowski, 1990].

where n_t is the length of the sequence tested for stationarity. If we divide T by $\sqrt{Var(T)}$, where $Var(T) = (n^3 s^2)/45$ [Schruben et al, 1983], we can justify treating

$$\hat{T} = \frac{\sqrt{45}}{n_t^{1.5} n_v^{0.5} \hat{s}\ [\overline{X}_{n_v}]} \sum_{k=1}^{n_t} k(1 - \frac{k}{n_t}) \left[\overline{X}_{n_t} - \overline{X}_k\right],$$

as at t statistic with d degrees of freedom[6], where $\hat{s}^2$ is the estimator of the variance s^2 and n_v is the length of the sequence used for estimating the steady state variance s^2. Heidelberger and Welch [1983] recommended n_v to be greater than 100.

A remaining problem in applying the test is that the steady state variance, s^2, is generally unknown. The sample variance, $\hat{s}^2$, can be used as the point estimator of the steady state variance which can be calculated over the latter portion of the collected data. This is done on the assumption that this latter portion of data is more representative of the steady state behaviour of the system, thus giving a better estimate of the steady state variance. For more information on the methods of obtaining the estimators of variance, see Fishman [1973], Ockerman and Goldsman [1997], and Goldsman and Tokol [2000].

The Schruben test for the presence of negative initialisation bias at a significance level of a is as follows:

Compute $\hat{s}^2$, and the test statistic, $\hat{T}$. Reject the hypothesis of no initialisation bias if $\hat{T} > t(d,\ a)$, where $t(d,\ a)$ is the upper $100a$-quantile of the t distribution with d degrees of freedom. If a test for positive initialisation bias is desired, then the sign of the test statistic $\hat{T}$ should be changed. A two-sided test for initialisation bias can be computed in the usual manner for t tests by using the absolute value of

[6] See pp. 289 of Fishman [1973] for the appropriate degrees of freedom for each estimator of s^2.

the test statistic $\hat{T}$. More details on the Schruben test can be found in Pawlikowski [1990] and Stacey [1993].

Figure 10 shows how the sequential Schruben test for detecting the length of the initial transient period has been implemented in Akaroa2. The observations to be tested for stationarity are within the test window. The variance window is used for estimating the steady state variance and it contains the collected data in the latter portion of the test window. If the observations in the test window pass the stationarity test, then the length of the initial transient period has been found, and we can assume that steady state has been reached. Otherwise, depending on the step length (see Figure 10), more observations need to be collected and the stationarity test is repeated over the new window of observations.

As expected, choosing different sizes for the test window, variance window and the step length can affect the performance of the test. The sequential version of the Schruben test implemented in Akaroa2 uses a heuristic for the first estimation of the length of the initial transient period. This heuristic was proposed by Gafarian [1978] and is described in detail in Pawlikowski [1990]. Using this heuristic, the length of the initial transient period is over after n_0 observations, if the time series $X_1, X_2, \ldots, X_{n_0}$ crosses the mean $\overline{X}(n_0)$ 25 times [Pawlikowski, 1990]. The step length is then estimated by half the length of the initial transient period found by this heuristic. The sizes of the variance window and the test window are initially fixed to 100 and 200. If the step length is larger than the test window then the size of the test window is changed to the step length. We will investigate the performance of the Schruben test in Chapter 3.

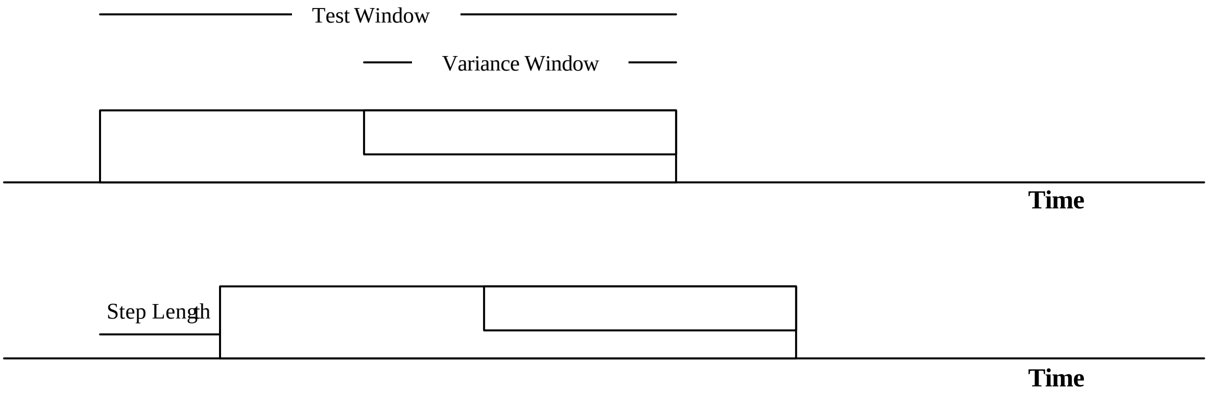

Figure 10. Sequential procedure for detecting the length of initial transient period used in Schruben test Implemented in Akaroa2.

2.2.2. A Modified Sequential Version of the Schruben Test

As it was discussed in Section 2.2.1, the Schruben test needs knowledge of the steady state variance. This could be accomplished by estimating the steady state variance over the latter portion of the collected data. However, this data still may not represent the steady state period. Stacey [1993] suggested that by moving the variance estimate window ten times the window size into the future from the test window (which contains the data to be tested for the initial bias), the problem would be solved only if the sizes of both windows are increased. He showed that the best results could be established with the test window size of $3000 \times r$ for the M/M/1/8 queue where r is the system load. Figure 11 shows this. We implemented the modification of the sequential version of Schruben test by moving the variance window 1000, 10000, 20000, and 50000 observations in future. The results are shown in Chapter 3.

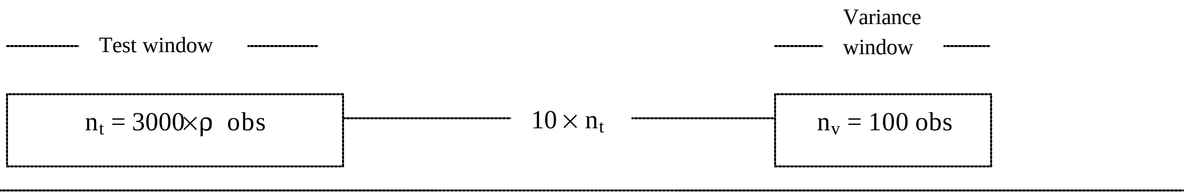

Figure 11. Sequential Procedure for Detecting the Length of Initial Transient Period used in the Modified Version of Schruben Test by Stacey.

2.2.3. GSS Tests

Goldsman, Schruben, and Swain [1991] introduced the following group of tests to detect the presence of a transient mean in a simulation process by comparing the variance estimators from different parts of a simulation run. These tests are based on the methods of batch means and standardized time series and their evaluation can be found in the paper by Cash, Nelson, Dippold, Long, and Pollard [1992]. All of these tests are based on an F statistic that compares the variability in the first portion of the output process to the variability in the latter portion of the process. Let $X_1, X_2, X_3, \ldots, X_n$ be a simulation output process in time-dependent order, and let $\overline{X}_n$ be the sample mean, a point estimate for the steady state mean, $E(X)$. For all the test statistics, partition $X_1, X_2, X_3, \ldots, X_n$ into b non-overlapping batches of m observations such that $n = b \times m$, and suppose the b batches are partitioned into two not necessarily equal-sized groups consisting of the first b' batches and the last b-b' batches (See Figure 12).

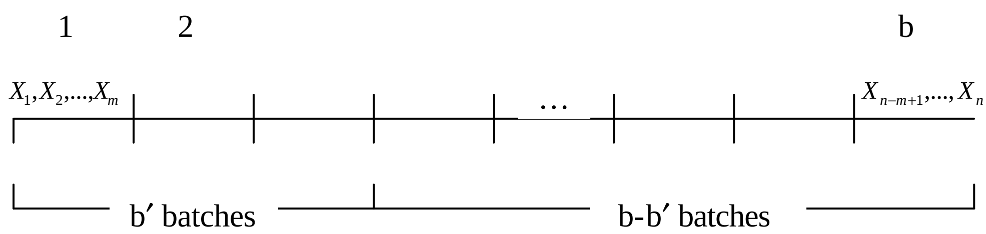

Figure 12. Partitioning Strategy used in GSS Tests.

Now, let V_{1st} be the variance estimator of the first b' batches and V_{2nd} be the variance estimator of the b-b' batches. Under the null hypothesis, the ratio $F = V_{1st} / V_{2nd}$ converges in distribution to an F random variable. The null hypothesis of no initial transient bias is rejected if $F > F_{1-\alpha,\ c,\ d}$ with the 1-a

quantile of an F distribution with c and d degrees of freedom (c and d varies for each statistical test and are defined in the following sections). These variance estimators of the first and second portions and the particular F distributions for each test are described in the Sections 2.2.3.1-2.2.3.3.

The batching strategy, which includes the total number of batches, b, and the fraction $f = b/b'$ clearly affects the power of the tests[7]. Cash et al [1992] evaluated the performance of the tests (discussed in further detail later in this section) with three different values of f and five different combinations of b and b' as listed in Table 6. They recommended b = 16 as the maximum number of batches.

b	**b′**	¦
2	1	0.5
8	4	0.5
16	4	0.25
16	8	0.5
16	12	0.75

Table 6. Batching Strategies used in Cash et al [1992].

To increase the chance that there is little bias in the second portion of the output process, Cash et al [1992] recommended f = 0.75. However, if the hypothesis of no bias is rejected then it is not clear how much data to delete. They discussed that rejecting the hypothesis when f = 0.75 does not mean that 75% of the data must be discarded. The strategy that they suggested is as follows: First perform the test with f = 0.25; if the null hypothesis is rejected, delete the first 25% of the data and apply the test again to the remaining data. If the null hypothesis is accepted,

[7]"For any statistical hypothesis test, there are two types of statistical errors, namely the error of rejecting the null hypothesis H_0 when it is true (Type I error) and the error of accepting H_0 when it is false (Type II error). The significance level of a test is the probability of making Type I error. The power of a statistical test is defined as the probability of not making Type II error, in other words, it is the probability of rejecting the null hypothesis H_0 when it is false" [Ma and Kochhar, 1993].

retest at $f = 0.5$ (and next at $f = 0.75$). The retest is needed because the hypothesis may be accepted when there is a significant bias in both the first and second portion of the process. Figure 13 shows an implementation of this strategy in the sequential analysis (described in Figure 10). We used this strategy in the implementation of the GSS tests.

The sequential version of the GSS tests requires that the batches b be partitioned into multiples of 4 (as f is equal to 0.25, 0.5, and 0.75). We used the batching strategy listed in Table 7 in our implementation of the GSS tests. Choosing a test window size n_t of 200 ($n_t = b \times m$) gives an appropriate number of observations in each batch, m, as listed in Table 7.

b	**m**
4	50
8	25
16	13

Table 7. Batching Strategies used in the implementation of GSS tests in Akaroa2.

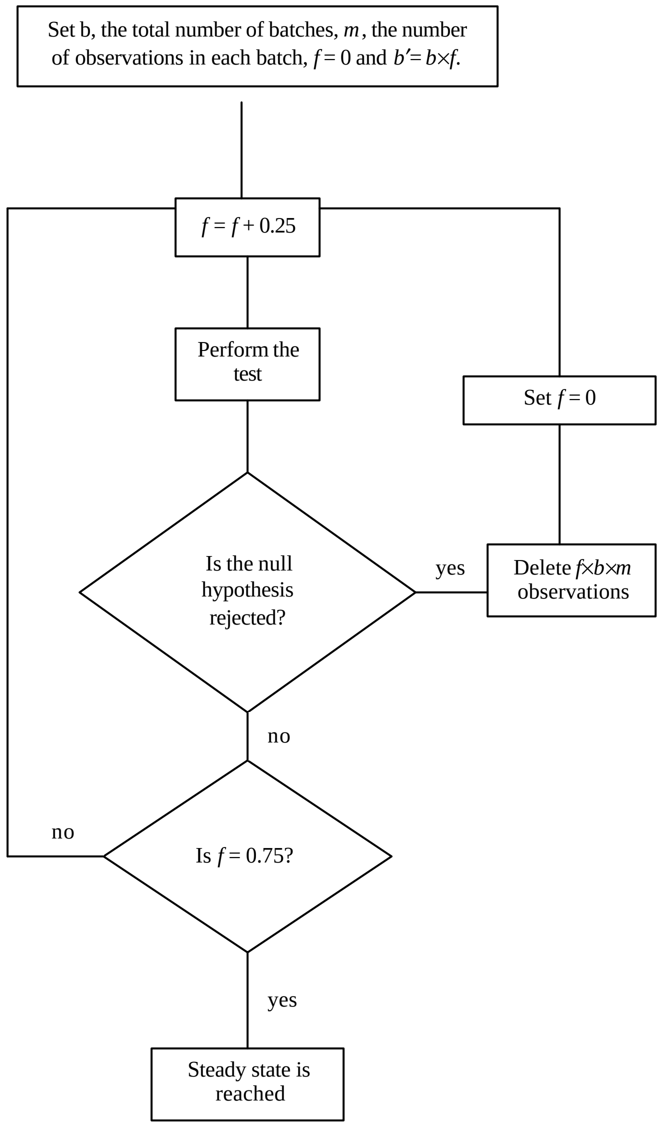

Figure 13. Sequential Procedure for Detecting the Length of the Initial Transient Period Based on the GSS Tests.

2.2.3.1. Batch-Means Test

This test is the simplest test of the group of tests proposed by Goldsman, Schruben, and Swain [1994]. It is based on the method of batch-means, in which the sequence of observations is divided into a series of batches. Let $i = 1, 2, ..., b''$ and $j = 1, 2, ..., m$, where b'' is the number of batches equal to b' or b-b' and m is the number of observations in each batch (see Figure 12).

The sample mean for the *i*th batch is

$$\overline{X}_i = \frac{1}{m}\sum_{j=1}^{m} X_{(i-1)m+j}$$

The variance estimator can be calculated as

$$V_{BM} = \frac{Q_{BM}}{b''-1}$$

where

$$Q_{BM} = m\sum_{i=1}^{b''}\left[\overline{X}_i - \frac{1}{b''}\sum_{j=1}^{b''}\overline{X}_j\right]^2$$

The critical value for this test is $F_{1-\alpha, b'-1, b-b'-1}$ [Cash et al, 1992].

2.2.3.2. Maximum Test

This test is based on the location and magnitude of the maximum deviation of the batch means. The test is for the presence of negative bias; an analogous test is available for the presence of positive bias. Let $i = 1, 2, ..., b''$ and $j = 1, 2, ..., m$, where b'' is the number of batches equal to b' or b-b' and m is the number of observations in each batch.

The sample mean for the ith batch is

$$\overline{X}_i = \frac{1}{m}\sum_{j=1}^{m} X_{(i-1)m+j}$$

and the running mean of all the batches is

$$\overline{X}_{i,j} = \frac{1}{j}\sum_{k=1}^{j} \overline{X}_{(i-1)m+k}$$

$$S_{i,j} = \overline{X}_i - \overline{X}_{i,j}$$

$$\hat{K}_i = \arg\max_{1 \le j \le m}\{jS_{i,j}\}$$

where argmax returns the index of a maximal element of the list.

$$\hat{S}_i = \hat{K}_i S_{i,\hat{K}_i}$$

$$Q_{MAX} = \sum_{i=1}^{b''} \frac{m\hat{S}_i^2}{\hat{K}_i(m-\hat{K}_i)}$$

and finally the variance estimator can be calculated as

$$V_{MAX} = \frac{Q_{MAX}}{3b''}$$

The critical value for this test is $F_{1-\alpha,\, 3b',\, 3b-3b'}$ [Cash et al, 1992].

2.2.3.3. Area Test

This test is based on the transformation of the data into a standardized time series and a computation of a variance estimator based on the area under the standardized time series. Let $i = 1, 2, ..., b''$ and $j = 1, 2, ..., m$, where b'' is the number of batches equal to *b´* or *b-b´* and m is the number of observations in each batch.

The running mean of all the batches is calculated by

$$\overline{X}_{i,j} = \frac{1}{j}\sum_{k=1}^{j} \overline{X}_{(i-1)m+k}$$

and the standardized time series *T* is defined as

$$T_{i,m}(t) = \frac{\lfloor mt \rfloor (\overline{X}_{i,m} - \overline{X}_{i,\lfloor mt \rfloor})}{s\sqrt{m}}$$

where $\lfloor . \rfloor$ is the greatest integer function. The area under the standardized time series can be calculated as,

$$\hat{A}_i = \frac{s}{m}\sum_{i=1}^{m} \sqrt{12} T_{i,m}(\frac{j}{m})$$

s^2 is the variance of the output process, which can be calculated using the same methods as in the Schruben test.

$$Q_{AREA} = \sum_{i=1}^{b''} \hat{A}_i^2$$

Then, the variance estimator can be calculated as

$$V_{AREA} = \frac{Q_{AREA}}{b''}$$

The critical value for this test is $F_{1-a,\, b',\, b-b'}$ [Cash et al, 1992].

2.2.3.4. Combined Tests

"Because these tests are asymptotically independent, we can combine the Batch-means test statistics with the Area and Maximum test statistics to create two more F-tests" [Cash et al, 1992]:

$$V_{BM+AREA} = \frac{Q_{BM} + Q_{AREA}}{2b'' - 1}$$

$$V_{BM+MAX} = \frac{Q_{BM} + Q_{MAX}}{4b'' - 1}$$

where b'' is the number of batches equal to b' or b-b'. The critical values for these two tests are $F_{1-a,\, 2b'-1,\, 2b-2b'-1}$ and $F_{1-a,\, 4b'-1,\, 4b-4b'-1}$ respectively [Cash et al, 1992].

The most common way for comparing tests is comparison of their power (see Cash et al [1992]). In the case of tests for detecting the initialisation bias, this can be done by introducing artificially generated stochastic sequences as input data to the tests and comparing the power of the tests to detect the bias. Let $X_{i,j} = Y_{i,j} - E(X)\, a_t$ where $X_{i,j} \equiv X_{(i-1)m+j}$ denote the jth observation from the ith batch ($i = 1,2,\ldots,b$ and $j = 1,2,\ldots,m$), $Y_{i,j}$'s be stationary observations with mean $E(X)$, and a_t's be the bias functions. Cash et al [1992] compared the power of the GSS tests using four bias functions. In all cases, they used 2400 input observations where half of them were contaminated by bias. The four bias functions are listed below:

1. Mean-shift bias function:

$$a_t = \begin{cases} 2r & t = 1,2,\ldots,1200 \\ 0 & t = 1201,\ldots 2400 \end{cases}$$

where r was chosen as 0.01, 0.1, and 0.25.

2. Linear bias function:

$$a_t = \begin{cases} 4r(1 - \frac{t}{1200}) & t = 1,2,\ldots,1200 \\ 0 & t = 1201,\ldots 2400 \end{cases}$$

where r was chosen as 0.01, 0.1, and 0.25.

3. Quadratic bias function:

$$a_t = \begin{cases} 6r(1 - \frac{t}{1200})^2 & t = 1,2,\ldots,1200 \\ 0 & t = 1201,\ldots 2400 \end{cases}$$

where r was chosen as 0.01, 0.1, and 0.25.

4. Damped, oscillating bias function:

$$a_t = \begin{cases} \frac{1614r}{p}\sin(\frac{pt}{300}) & t = 1,2,\ldots,1200 \\ 0 & t = 1201,\ldots 2400 \end{cases}$$

where r was chosen as 0.01, 0.1, and 0.25. The value used for p was not specified by Cash et al [1992] in this function.

They showed that the Maximum test was the most powerful, while the Batch-mean and the Area tests were the least powerful. We will compare the performance of the GSS tests in Chapter 3.

2.2.4. Rank Test

The Rank test was proposed by Vassilacopoulos [1989] to detect the presence of the initialisation bias based on the properties of a rank statistics. Suppose that the performance measure of interest of a simulation system is the mean, *E(X)*, of a stochastic process. Let $X_1, X_2, X_3, ..., X_n$ be the observations to be tested for initialisation bias with respect to *E(X)*. The null and alternative hypotheses take one of the following forms:

1. H_0 : No positive initialisation bias exists in the observation sequence.
   H_1 : Positive initialisation bias exists in the observation sequence.

2. H_0 : No negative initialisation bias exists in the observation sequence.
   H_1 : Negative initialisation bias exists in the observation sequence.

3. H_0 : No initialisation bias exists in the observation sequence.
   H_1 : Initialisation bias exists in the observation sequence.

The procedure for the Rank test is summarized as follows:

1. Find the ranks of the observations $X_1, X_2, X_3, ..., X_n$. That is, a new sequence $R_1, R_2, R_3, ..., R_n$ is obtained such that R_i is the ascending rank of X_i. The functional definition of the rank R_i of observation X_i in a set of n observations is given by

   $$R_i = \sum_{j=1}^{n} S(X_i - X_j)\ ,\ i = 1,2,...,n\ ,$$

   where

$$S(X_i - X_j) = \begin{cases} 1, & if\,(X_i - X_j) \geq 0, \\ 0, & otherwise. \end{cases}$$

Subsequently, compute a new sequence $U_1, U_2, U_3, ..., U_n$ based on the following equation:

$$U_i = U_{i-1} + 2R_i - (n+1), \quad i = 1, 2, ..., n \,,$$

where U_0=0.

2. Compute the following test statistics:

$$c^+ = \max \{U_i,\, i = 1,2,...,n\}\,,$$
$$c^- = \min \{U_i,\, i = 1,2,...,n\},$$
$$c = \max \{|\,U_i\,|,\, i = 1,2,...,n\}\,,$$

and the following significance levels associated with c^+, c^- and c, which were calculated based on the central limit theorem (see Vassilacopoulos [1989] for more detail):

$$\hat{a}^+ = \exp[\frac{-6(c^+)^2}{(n^3 + n^2)}]\,,$$
$$\hat{a}^- = \exp[\frac{-6(c^-)^2}{(n^3 + n^2)}]\,,$$
$$\hat{a} = 2\exp[\frac{-6c^2}{(n^3 + n^2)}]\,.$$

3. Let a be the selected significance level of the test. Then,

- If $\hat{a}^+ < a$, reject the hypothesis of no positive bias, or
- If $\hat{a}^- < a$, reject the hypothesis of no negative bias.

When the direction of initialisation bias is uncertain, a two-sided test is applicable. That is:

- If $\hat{a} < a$, reject the hypothesis of no negative bias.

An evaluation of the performance of the Rank test and the Schruben test by comparing their abilities to detect the initialisation bias in so-called contaminated stochastic processes was performed by Ma and Kochhar [1993]. They found that the Schruben test was more sensitive than the Rank test, meaning that when initialisation bias did exist, the optimal test had a higher power to detect the initialisation bias than the Rank test. However, when there was no initialisation bias, the Rank test performed better. Overall, they concluded that the Schruben test performed better than the Rank test. However, the Rank test was simpler to implement for two reasons. First, it did not require the estimation of the variance of a given stochastic sequence. Second, an efficient ranking algorithm, that was required for its application was easily available.

2.2.5. LWS Test

Lada, Wilson, and Steiger [2003] described a procedure for detecting the length of the initial transient period. Using this procedure, a batch size and an initial transient period based on the independence of the batches were determined. The procedure for the LWS test can be summarized as follows:

Divide the initial sample $X_1, X_2, X_3, ..., X_n$ into b = 256 adjacent batches of size m = 16 and define the initial spacer S ,which is the number of ignored batches between the remaining batches, equal to 0.

The *i*th batch mean is defined as,

$$\overline{X}_i = \frac{1}{m}\sum_{j=1}^{m} X_{(i-1)m+j}$$

and the grand average of the b batch means as,

$$\overline{X}_{m,b} = \frac{1}{b}\sum_{i=1}^{b} \overline{X}_i$$

The null hypothesis is that the $\overline{X}_i$, i = 1, 2, …, b , are independent, and identically distributed. Using the Von Neumann randomness [1941], the test statistic C_b is defined as the ratio of the mean square successive difference to the variance,

$$C_b = 1 - \frac{\sum_{j=1}^{b-1}(\overline{X}_j - \overline{X}_{j+1})^2}{2\sum_{j=1}^{b}(\overline{X}_j - \overline{X}_{m,b})^2}$$

The null hypothesis is accepted if $|C_b| \le z_{1-a/2}\sqrt{\frac{b-2}{b^2-1}}$

If the null hypothesis for b = 256 is rejected, then we reduce b to 128 batches by inserting spacers $S = m$, consisting of one ignored batch between each 128 remaining batches. That is every other batch mean, beginning with the second batch mean is retained, and the alternate batch means are ignored. Again, the remaining 128 batches are tested for randomness and if the batch means pass the randomness test then the length of the initial transient period is detected as S. Otherwise b is reduced to 85 batches by inserting spacers $S = S + m$, consisting of two ignored batches between each 85 remaining batches. This process is continued until either the length of the initial transient is detected or b is reduced to 25. At this point the procedure is repeated by resetting m to $\lfloor \sqrt{2m} \rfloor$, b = 256, and S = 0 [Lada, Wilson and Steiger, 2003]. Figure 14 describes this procedure. Performance results of the LWS test will be examined in Chapter 3.

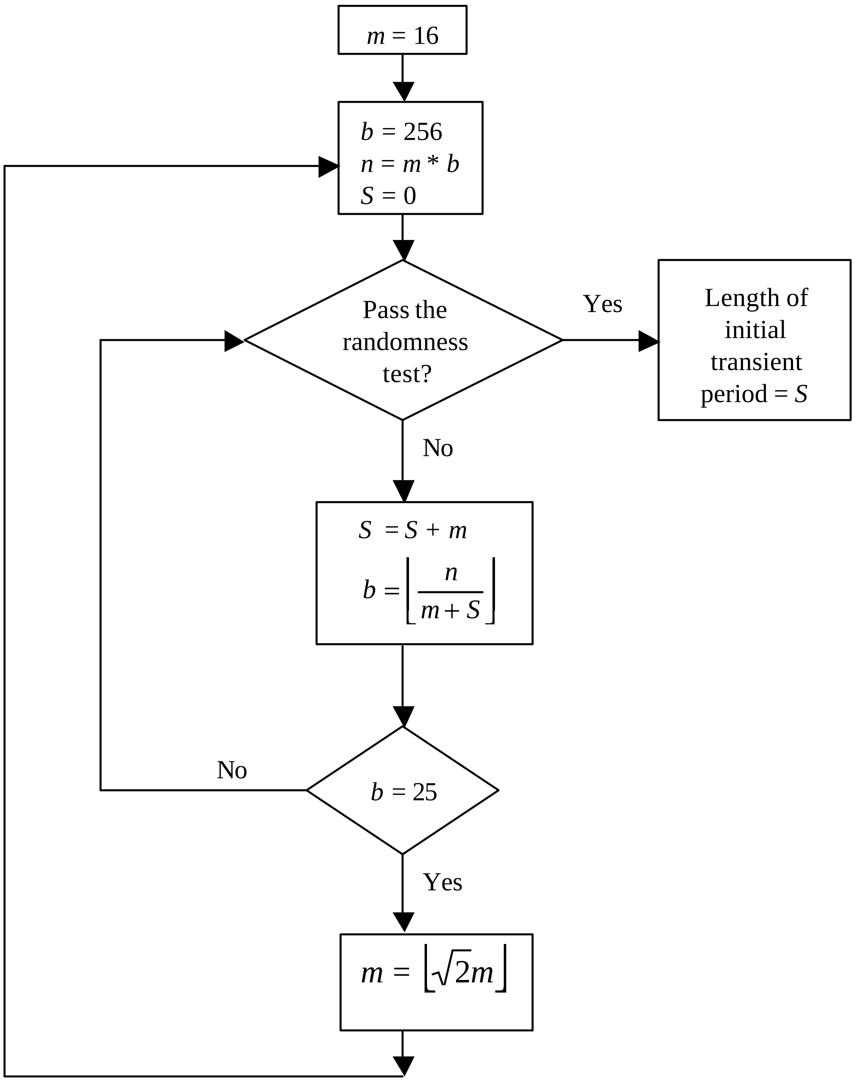

Figure 14. Flowchart of the LWS Test.

2.2.6. Yücesan Test

Yücesan [1993] proposed a test to detect the length of the initial transient period based on the independence of the batches. The following explains the Yücesan test, as described in Stacey [1993]:

“The collected observations are divided into b batches of size m. The batches are then split into two groups and the mean of each batch is calculated. The first group initially contains the first batch mean and the second group, the other b-1 batch means. The means within each group are then averaged to create group means. These two group means are then compared and the actual test statistics is calculated by the absolute difference of these two group means. The order of all b batch means is then randomised and split in the same way to compute a statistics. This statistics is calculated for N_s (Yücesan used N_s = 99, 199, and 999 number of shuffles) different randomised ordering of the batches. The significance level of the test is calculated from the number of times the statistics is greater than or equal to the actual statistics. If the significance level has reached the desired value, the procedure is finished and the initial transient period is said to be over at the end of the first group of unshuffled batches. If the significance level has not been reached, we retest the hypothesis through the randomisation test by moving the first batch mean from the original ordering of the second group to the first group (so the first group contains 2 means while the second contains b-2 means) and the whole process is repeated”. For more detail, see Figure 7 of Stacey [1993].

Stacey [1993] compared the length of the initial transient detected by the Schruben test with the one detected by the Yücesan test. He concluded that the Yücesan test detected a length closer to the one obtained by the relaxation time compared to the Schruben test.

2.2.7. Initial Transient Detection Using Parallel Replications

All the previous tests were based on the analysis of the simulation output data from a single simulation run for detection of the presence of the transient effects in mean values, leaving open the problem of the initial transient in analysis of the other statistics, such as variances, or probability distribution. A procedure that can detect the presence of bias in steady state analysis of probability distribution provides information on the initialisation bias in the mean, variance, and other performance measures. However, a test that can detect the initialisation bias in steady state probability distributions will detect a longer length of initial transient period than a test that can detect the bias of the mean.

In the method of independent replications, each independent replication starts with a different seed, so each replication may have a different length of the initial transient period. Therefore, the first l observations from each replication are deleted. This result is not very satisfactory, as for a large number of replications a large number of observations need to be deleted. Bause and Eickhoff [2002] proposed a procedure to detect the presence of the initial transient using parallel replications based on the probability distribution of the observations (convergence of the cumulative distribution function (CDF) towards the steady state distribution). This is the summary of the procedure:

We split the model time into Time Intervals (TI_j's) of length t_i, where $j = 1, \ldots, n$ and n is the amount of observed intervals. Now let k denote the number of replications and X_{ij} ($i = 1,\ldots, k$; $j = 1,\ldots, n$) be the transformed output of the replications. For example x_{ij} can be defined as the number of observations in time interval j at replication i, divided by the size of interval j. Using this definition, $X_{1j}, \ldots, X_{kj}$ is the random sample of the jth x-interval: RS_j. This is illustrated in Figure 15.

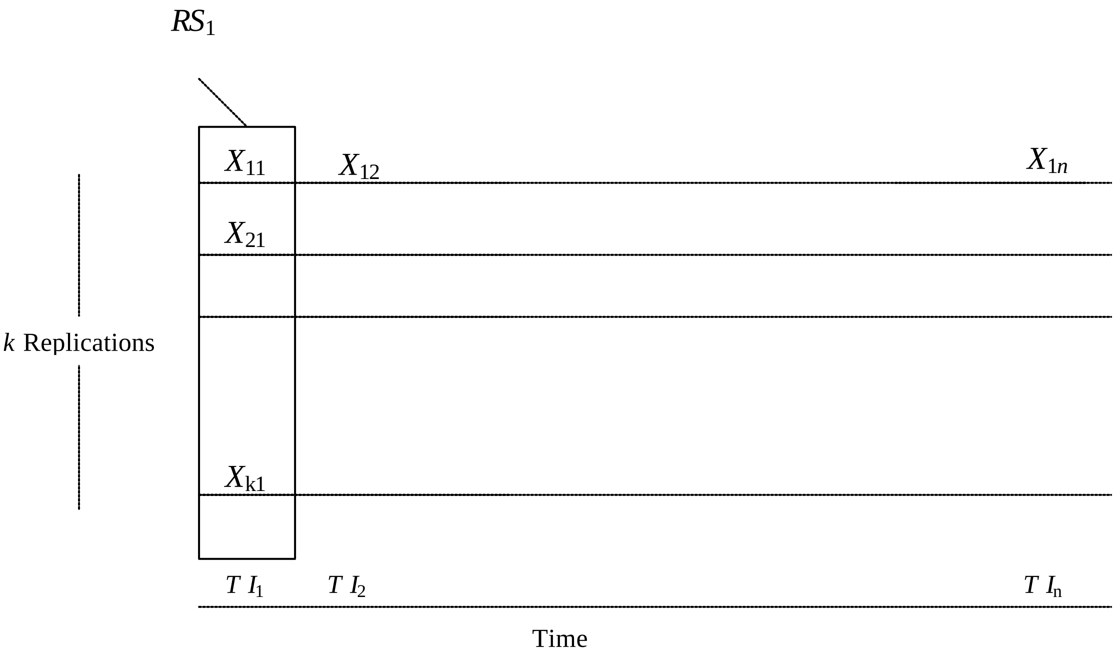

Figure 15. Procedure to Detect the Presence of the Initial Transient using Parallel Replications Based on the Probability Distribution of the Observations.

"In steady state, there would be no change in the probability distribution" [Bause, and Eickhoff, 2002]. To find the length of the initial transient period, the following procedure will check if all the random samples, RS_j, are realisations governed by the same probability distribution.

1. "Choose a ratio 1:r (a predefined ratio between the initial transient and the steady state period which was chosen as 1/10 by Bause and Eickhoff [2003]), and a safety-level $0 \leq p \leq 1$, to consider the error of the statistical method for random sample comparisons.
2. Initialise the test sample, $TS := RS_0$ and $n := 0$.
3. Observe $r + 1$ new intervals of all replications and compute the $r + 1$ new random samples: $RS_{n+1}, \ldots, RS_{n+r+1}$.
4. Increase the amount of data: $n := n + (r + 1)$.
5. Select a new test sample: $TS := RS_{\frac{n}{r+1}}$.

6. Compare TS with RS_j for $j = \frac{n}{r+1} + 1, \cdots, n$ (using the Kolmogorov-Smirnov test to compare two random samples). If more than $(p \times 100)\%$ of the compared random samples have a different probability distribution than TS: goto 3.
7. Calculate the initial transient length $l := \frac{n}{r+1}$" [Bause, and Eickhoff, 2003].

The performance of the statistical tests surveyed in this Chapter will be evaluated in Chapter 3.

Chapter 3

Comparison of the Performance of the Methods for the Detection of the Initial Transient

We looked at the survey of the statistical tests and the theoretical results on the relaxation time and the algorithm by Kelton and Law to detect the length of the initial transient period in Chapter 2. In this chapter, we compare performance of the following statistical tests: the Area test, the Maximum test, the Batch-mean test, the Schruben test, the Modified Schruben test, the LWS test, and the Rank test, with the relaxation time and the algorithm by Kelton and Law. We investigate two issues; the length of the initial transient period in the mean waiting time of the customers in the system detected by these tests compared to the theoretical measures, and the accuracy of these detections. In Section 3.1, we look at the implementation of these tests, the changes needed for these statistical tests to be compatible with Akaroa2[8], and the design issues for these experiments. We look at the results of these experiments in Section 3.2.

[8] Only the sequential version of the statistical tests can be used in Akaroa2.

3.1. Implementation of the Statistical Tests to Detect the Length of the Initial Transient Period

A sequential version of the Schruben test is implemented for detecting the length of the initial transient period in Akaroa2 as mentioned in Section 2.2.1. This sequential Schruben test uses two windows: the test window that contains the observations to be tested for stationarity and the variance window that contains the latter portion of the collected data in the test window for estimating the steady state variance (see Figure 10). The length of initial transient period is found when the observations in the test window pass the stationarity test. Otherwise, depending on the step length, more observations are collected and the stationarity test is repeated over the new set of observations.

As expected, choosing different sizes for the test window, the variance window, and the step length, can affect the performance of the test. The sequential version of the Schruben test implemented in Akaroa2 uses a heuristic to decide on an estimated length of the initial transient period. This heuristic was proposed by Gafarian [1978] and is described in detail in Pawlikowski [1990]. Using this heuristic, the length of initial transient period is taken as over after n_0 observations if the time series x_1, x_2, …, x_n crosses the mean, $\overline{x}(n_0)$ 25 times [Pawlikowski, 1990]. The step length is then estimated as half of the length of initial transient period found by this heuristic. The sizes of the variance window and the test window are initially fixed to 100 and 200 respectively. If the step length is larger than the test window then the size of the test window is changed to the step length.

Considering this approach, occasionally the length of the initial transient period found by the heuristic is very long. Thus, the number of observations to be tested in the test window, which is determined from the heuristic, is very large. In this case, a large number of observations in the test window are deleted and this sequential version of the Schruben test suggests a long initial transient period. For this reason, in addition to the Schruben test implemented in Akaroa2, we used a fixed window size for the implementation of the Schruben, the Modified Schruben, the Rank, and the GSS tests in Akaroa2.

In our implementation of the Schruben and the Modified Schruben tests, the same sequential method described in Section 2.2 was used and the sizes of the test window, the variance window, and the step length were set to 200, 100 and, 50 respectively. In the case of the Rank test, the window size and the step length were also set to 200 and 50 respectively. The window size was set to 200 for the GSS tests, but the step length was set to either 50, 100, or 150 following the sequential method described in Section 2.2.3.

3.1.1. Queueing Models

Three different queueing systems were chosen for the evaluation of the performance of the tests to detect the length of the initial transient period in the mean waiting time of the customers in the system:

1. M/M/1/8 with the coefficient of variation of the service time equal to 1.
2. M/$Erlang_4$/1/8 with the coefficient of variation of the service time equal to 0.5.
3. M/$Pareto_{a=2.1}$/1/8 with the coefficient of variation of the service time equal to 2.18.

An experiment on the detection of the length of initial transient period for the M/M/1/8 queue was first performed for a fixed number of independent runs set to 3000 using the Schruben test. The average result of these 3000 independent runs was then used as a single point estimator.

Figure 16 shows the performance of the Schruben test when measuring the length of the initial transient period of the mean waiting time of customers in an M/M/1/8 queueing system. As mentioned, for each system load, 3000 runs were collected and averaged. As each of these runs is independent, the detected length of the initial transient period for each run is independent from the others. The error bars on this graph indicate the variability of the length of the initial transient period detected by the Schruben test. As expected, Figure 16 shows that the detected lengths are more variable when the system is in a higher traffic load.

To solve this problem, instead of collecting a fixed number of runs, thus achieving a high error in the variance for different system load, a fixed absolute error was required in our research. This was done by checking the error of the collected results – the lengths of initial transient period – after each run. If this error was larger than a fixed value, more runs were collected, otherwise the process was stopped. The absolute error of the results for the system load 0.55 (equal to 7.8 after 3000 runs, equivalent to the relative error of 2.5%) was chosen as the acceptable value of the absolute error for all results[9]. A problem that arises with this sequential procedure is that as the error is very small at low traffic loads, the number of runs needed could be as small as one or two. For this reason, we selected a minimum number of 50 runs, even though fewer runs are needed to reach the fixed absolute error.

[9] The absolute error of 7.8 was obtained when executing 3000 runs at system load 0.55. This value was chosen as a reasonable trade off between collecting a small number of runs and low accuracy (by choosing the relative error of the results in a high system load) and collecting a large number of runs and higher accuracy (by choosing the relative error of the results in a low system load).

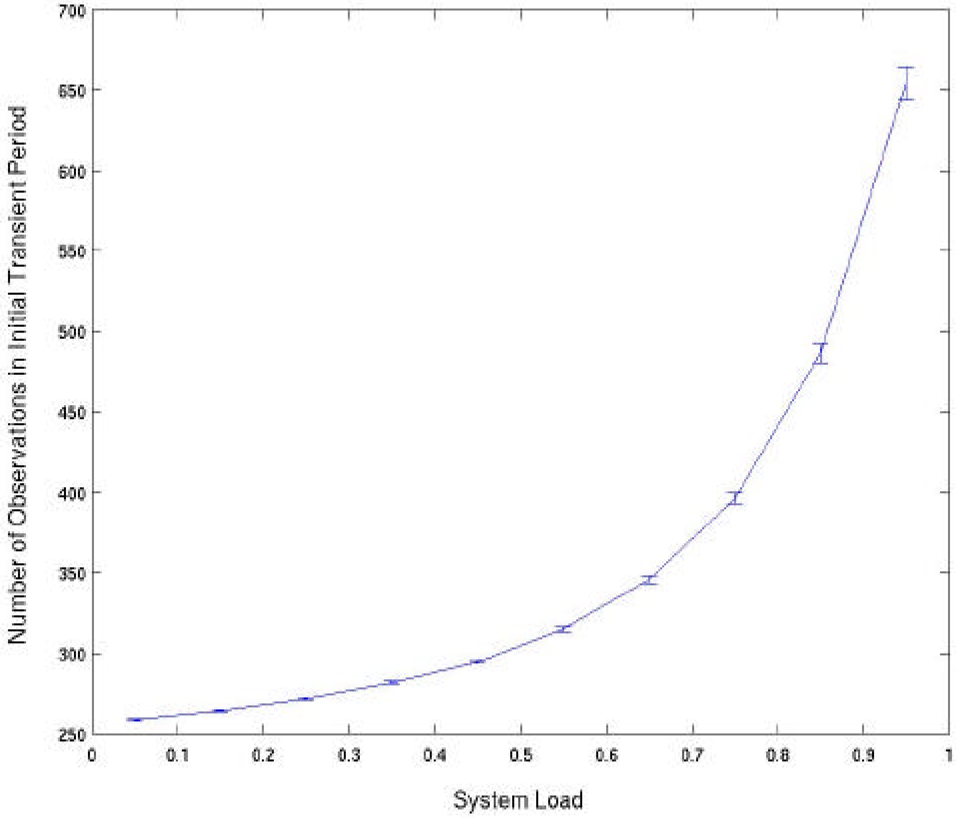

Figure 16. Performance of the Schruben Test after 3000 Runs for the M/M/1/8 Queue (with maximum relative error below 3%, at 0.95 confidence level).

3.1.2. Implementation of the GSS Tests

As mentioned in Section 2.2.3, Cash et al [1992] state that the batching strategy, which includes the total number of batches, b, and the fraction of b to b', f, clearly affects the power of the tests (see Section 2.2.3). Our implementation of the GSS tests was performed using three different batch numbers, b, and three different numbers of observations in each batch, m. Following the sequential method described in Figure 9, we used a window size of $b \times m \cong 200$ so that the GSS tests can be compared to the Schruben test and the Rank test with a window size of 200. We chose 16, 8, 4 and 13, 25, 50 for the values of b and m, respectively. The results of these experiments are discussed in Section 3.2.2.

3.1.3. Implementation of the Modified Schruben Test

In the Schruben test we need to know the steady state variance and this is usually obtained by estimating the variance over the latter portion of the collected data (described in more detail in Section 2.2.2). However, these data may still not represent steady state. To improve the steady state variance estimate, we can move the variance estimate window away from the test window into the future. The question that arises here is how far into the future we should move the variance window away to get the variance estimate more likely within the steady state period? We conducted an experiment for four different cases, moving the variance window 1000, 10000, 20000, and 50000 observations away from the test window. The results of this experiment are discussed in Section 3.2.3.

3.2. Evaluation of the Statistical Tests to Detect the Length of Initial Transient Period

3.2.1. Number of Runs Required for Obtaining a Fixed Error of the Results in Different System Loads

The number of runs needed to detect the length of the initial transient period was determined sequentially by using a fixed absolute error of the results as was described in Section 3.1. Table 8 shows the number of runs required to reach this fixed absolute error (equal to 7.8) of the detected lengths of the initial transient period for an M/M/1/8 queue using the procedure described in Section 3.1 with a significance level of $a = 0.05$. As can be seen, the Rank test needs the highest number of runs to reach the same absolute error. This suggests that within a set of independent runs, the Rank test produces more variable lengths of the initial transient period than the other tests. This table also shows that — except for the Rank test — the number of runs needed by the LWS test is higher than those required by the remaining tests. The number of runs for the Schruben test currently implemented in Akaroa2 (with variable window sizes) is also higher than the other versions of Schruben tests and the GSS tests. These results were expected as the number of observations in the test window was determined by a heuristic that would occasionally detect overly large lengths of the initial transient period. Comparing the rest of the tests in the table, at a higher traffic load, the Schruben, and Modified Schruben tests with a fixed window size need a slightly higher number of runs than the GSS tests.

Load \ Tests	Schruben Test in Akaroa2	Schruben Test with Fixed Window Length	Modified Schruben Test with Fixed Window Length	Batch-mean Test	Maximum Test	Area Test	LWS Test	Rank Test
0.05	50	50	50	50	50	50	50	50
0.15	50	50	50	50	50	50	50	50
0.25	50	50	50	50	50	50	50	50
0.35	50	50	50	50	50	50	50	50
0.45	50	50	50	50	50	50	50	261
0.55	50	50	50	50	50	50	50	846
0.65	104	50	61	54	51	50	50	3411
0.75	130	52	81	84	76	52	50	23024
0.85	1224	361	377	307	299	301	932	275228
0.95	18614	1102	1224	942	930	928	98295	500000+

Table 8. Table of the Number of Runs Needed to Reach a Fixed Absolute Error for the M/M/1/8 Queueing System.

3.2.2. Variation of the Length of Initial Transient Period

To observe the behaviour of these tests in more detail and clarify the results found in Table 8, we drew histograms of the length of initial transient period from 3000 runs for each of the tests in two different system loads, *?*: medium traffic and heavy load traffic. The results of these experiments are shown in Figures 17-22 for system load 0.55 and Figures 23-28 for system load 0.95 in the M/M/1/8 queueing system. These results confirm the results shown in Table 8 and show that the Rank test produces the most variable estimates of the length of initial transient period. Figure 22 and Figure 28 show the Rank test with 0.55 and 0.95 system load, respectively. Figure 28 also shows that for a heavy system load such as 0.95, the Rank test detects lengths as long as 200000 observations (much larger than the estimation by the relaxation time, see Section 3.2.5), thus making this test highly impractical. In Section 3.2.5, we will further investigate this overestimation by comparing the Rank test with the relaxation time and other proposed tests. Figures 21 and 27 represent the behaviour of the LWS test in the system load 0.55 and 0.95 respectively. Figure 21 shows in a low traffic system, the LWS test detect lengths not as variable as the Schruben and the GSS tests. The high number of runs needed to reach the absolute error of 7.8 for the system load 0.95 (see Table 8) can be clarified by looking at the Figure 27. As this figure shows, the LWS test detects an extremely variable length of initial transient period (additionally, the range of these lengths is 10 times longer than the GSS and the Schruben tests). The gaps between the bars in Figure 27 can be explained as the LWS test detects only lengths multiples of m (m = 16, $\sqrt{2 \times 16}$, …). The histograms also show that except for the LWS test, the other tests detect a length of initial transient period over 200. This is because as we used the sequential method described in Figure 9 for the implementation of these tests, the minimum length of the initial transient period is equal to the window size, which is set at 200.

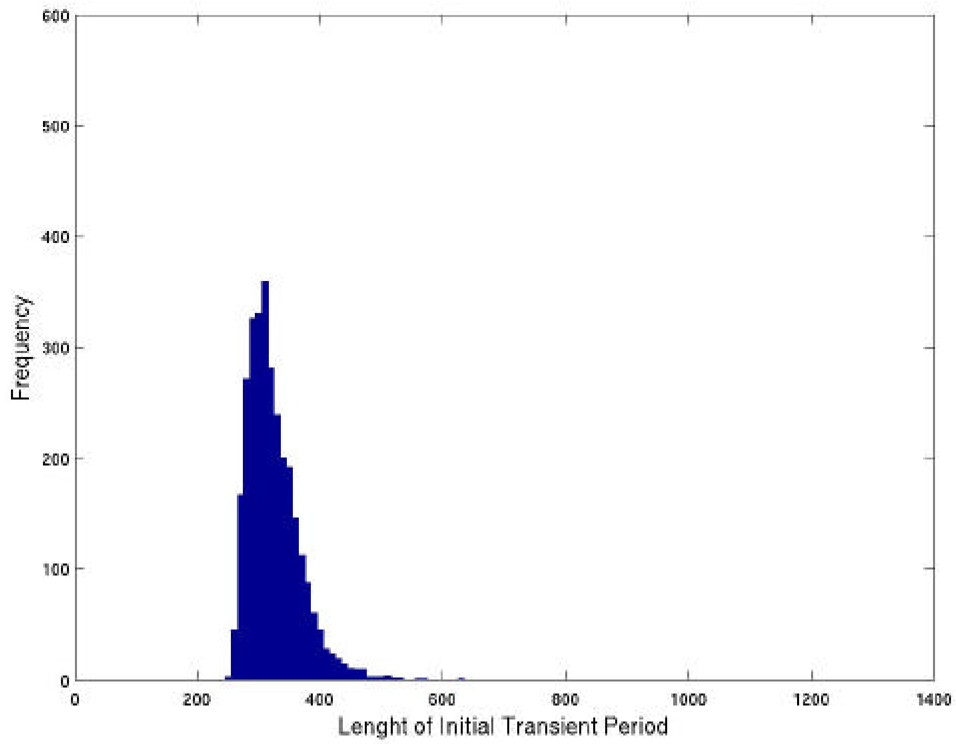

Figure 17. Histogram of Length of Initial Transient Period using the Area Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.55.

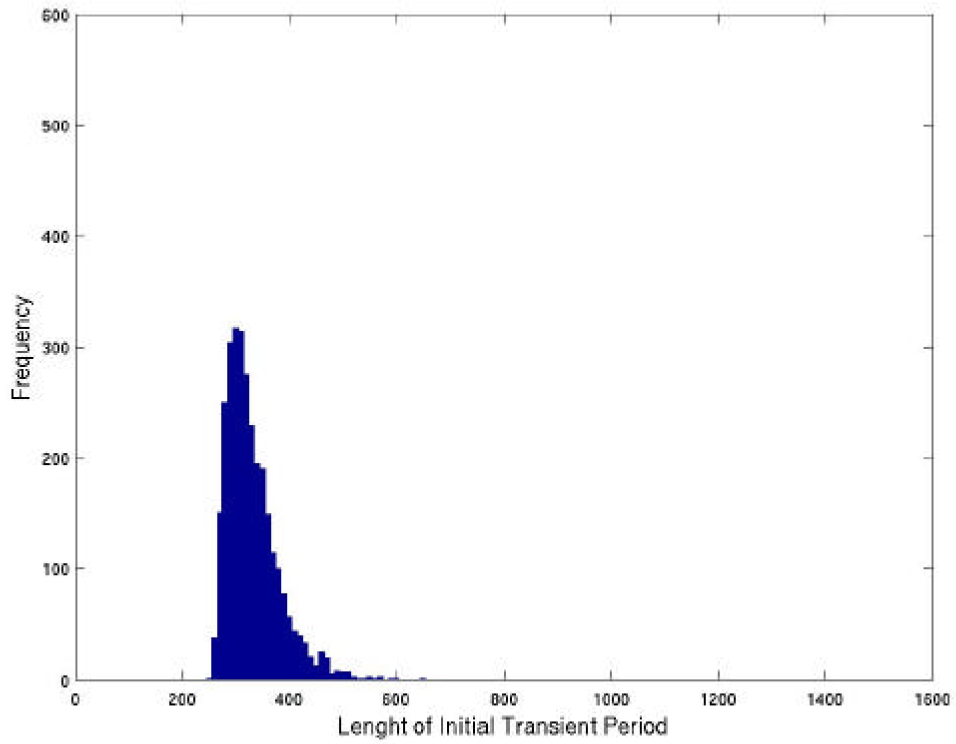

Figure 18. Histogram of Length of Initial Transient Period using the Batch-mean test with 3000 Runs for the M/M/1/8 Queue at ? = 0.55.

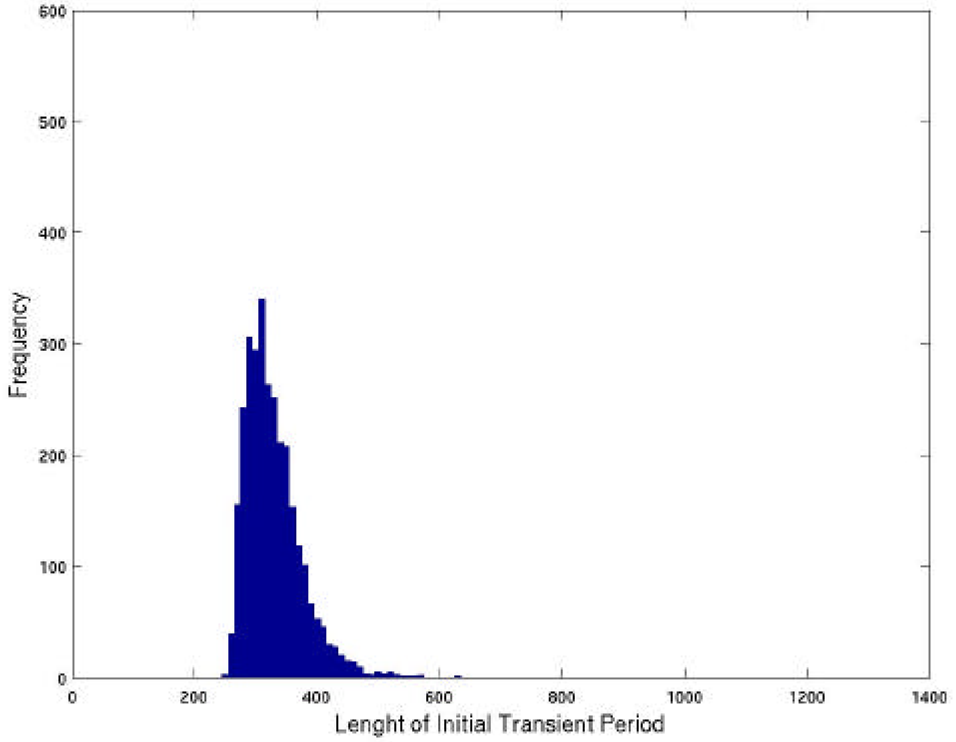

Figure 19. Histogram of Length of Initial Transient Period using the Maximum Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.55.

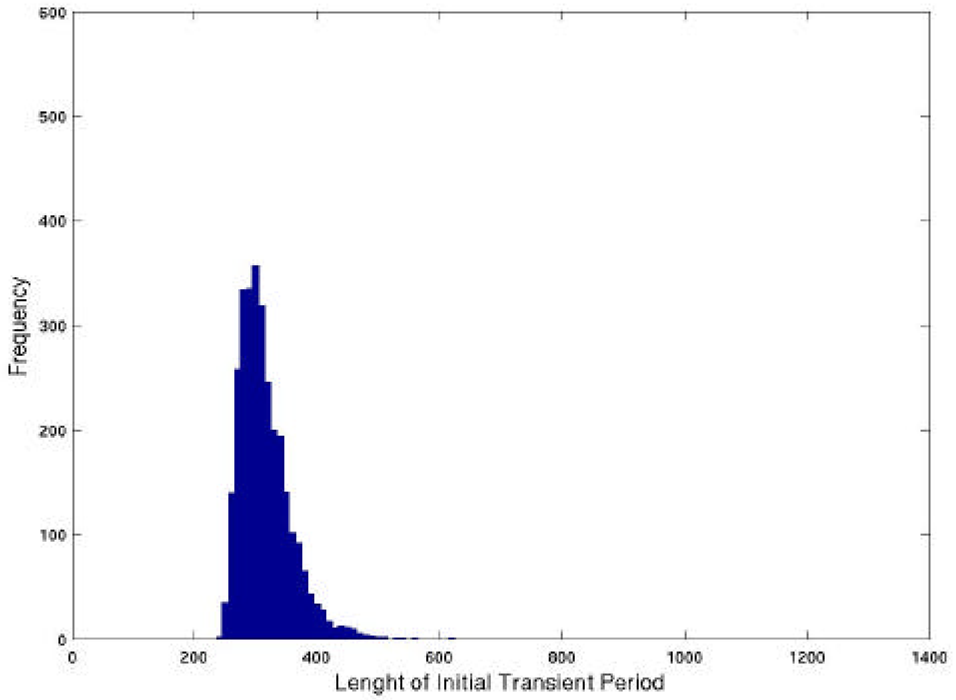

Figure 20. Histogram of Length of Initial Transient Period using the Schruben Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.55.

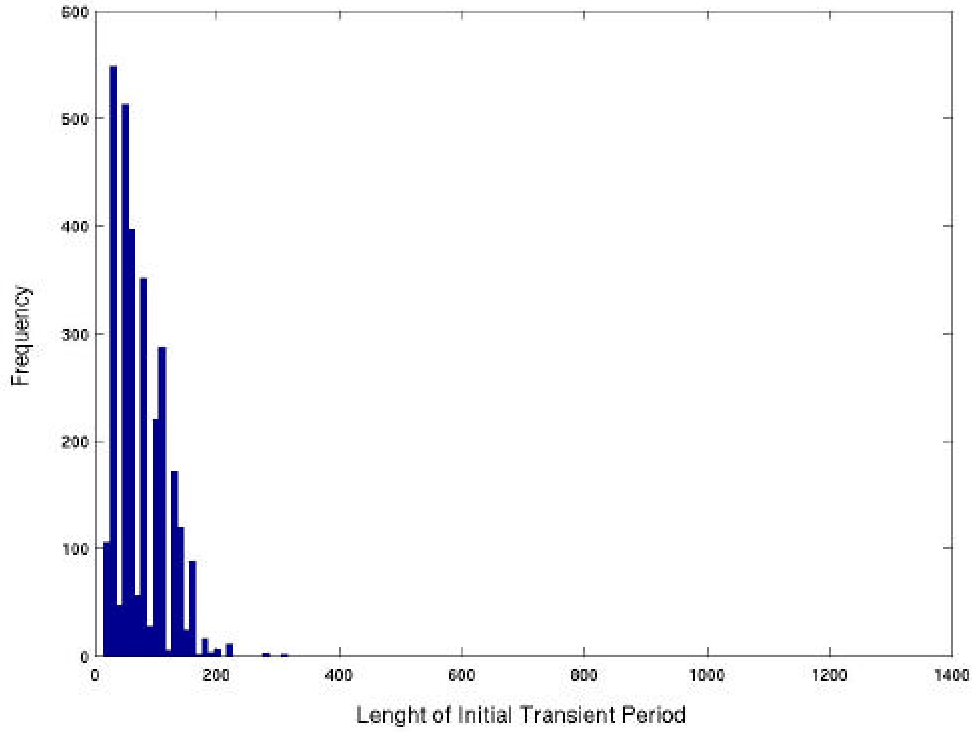

Figure 21. Histogram of Length of Initial Transient Period using the LWS Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.55.

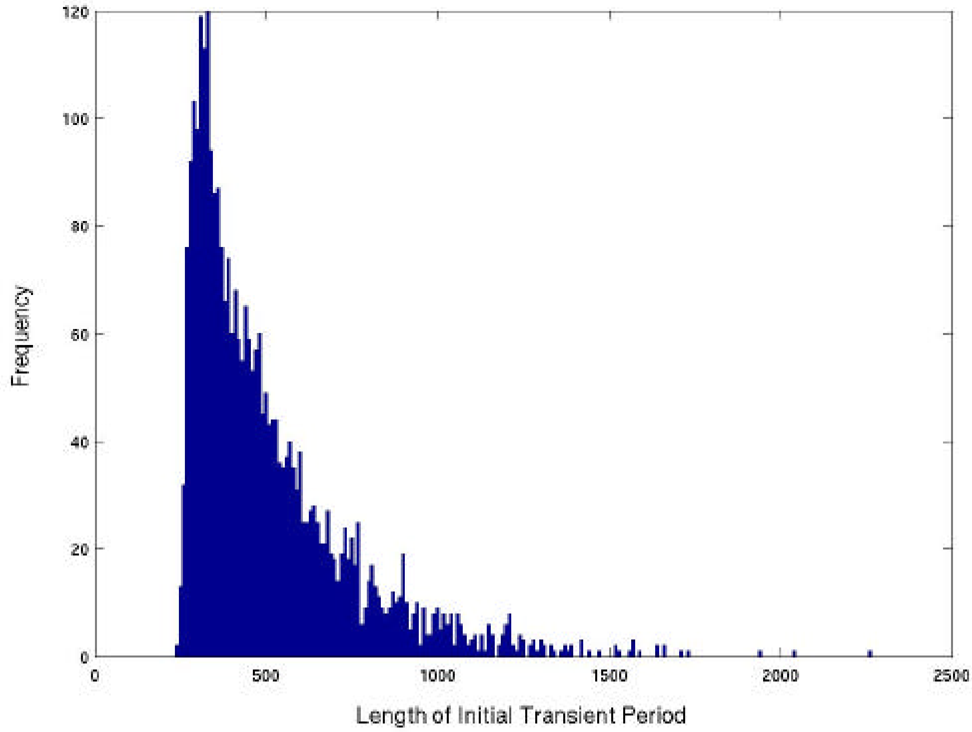

Figure 22. Histogram of Length of Initial Transient Period using the Rank Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.55.

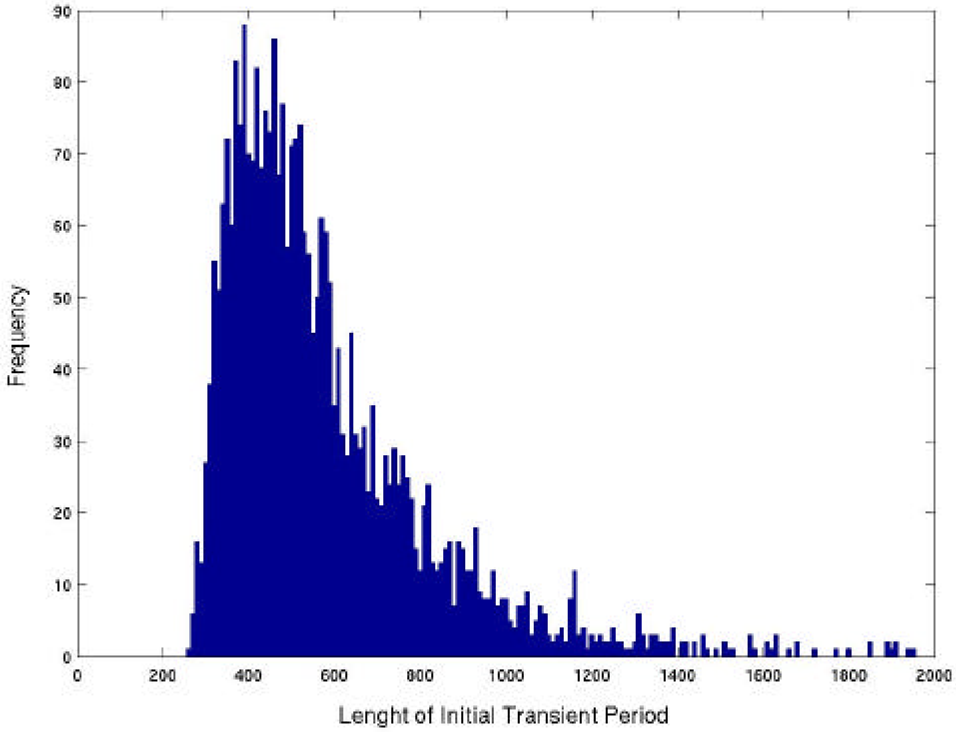

Figure 23. Histogram of Length of Initial Transient Period using the Area Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.95.

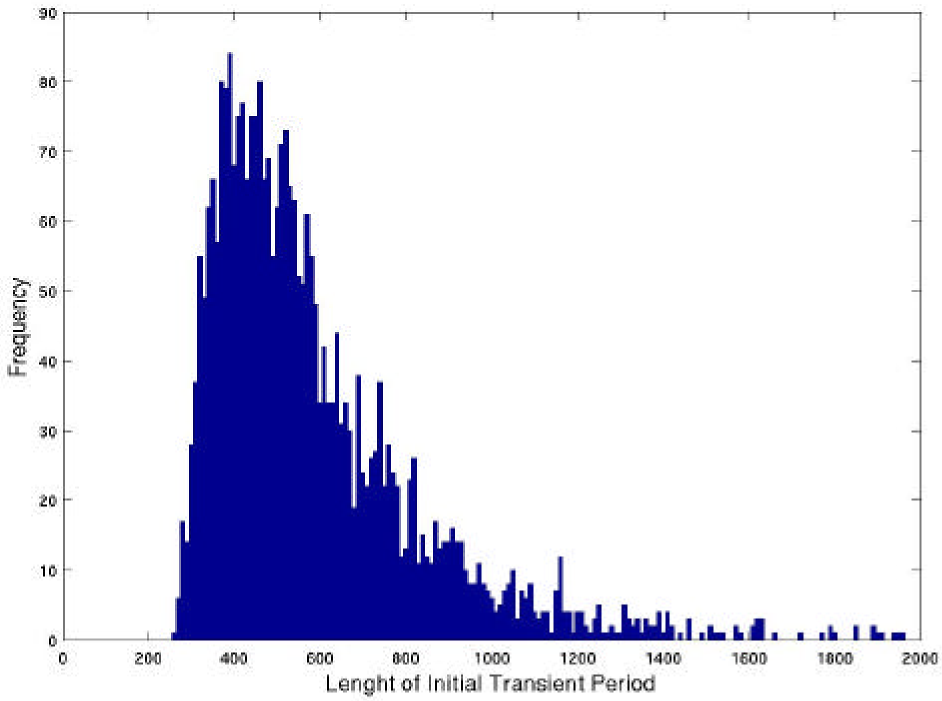

Figure 24. Histogram of Length of Initial Transient Period using the Batch-mean test with 3000 Runs for the M/M/1/8 Queue at ? = 0.95.

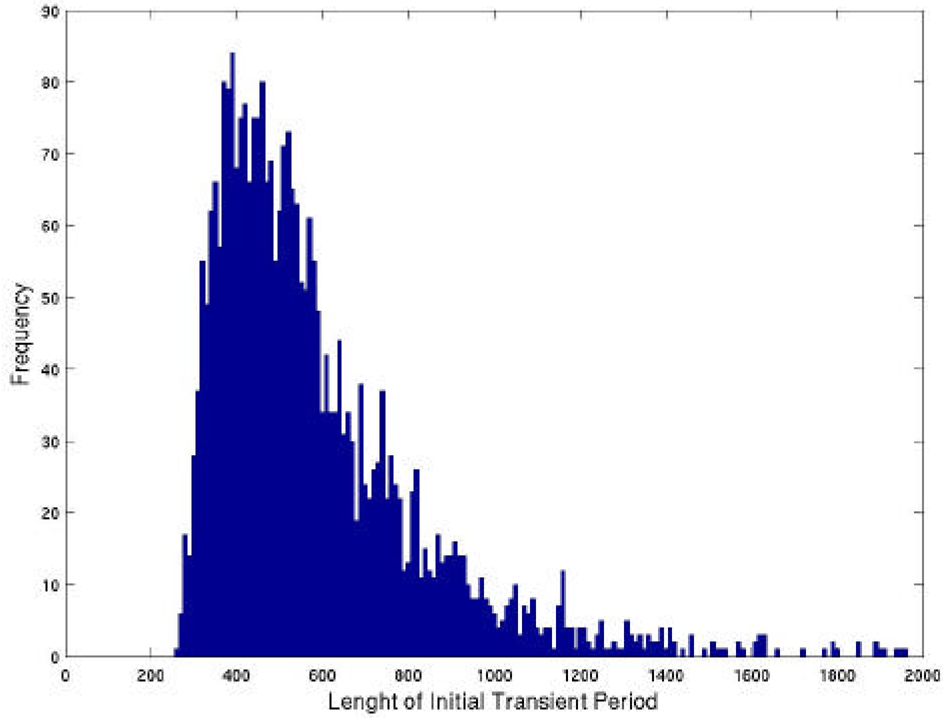

Figure 25. Histogram of Length of Initial Transient Period using the Maximum Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.95.

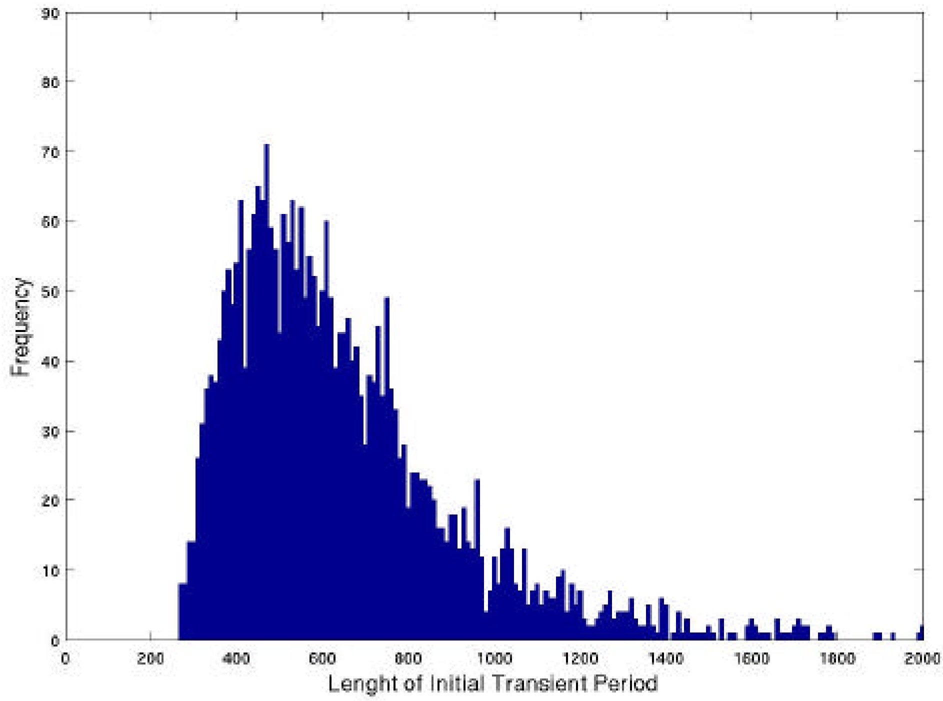

Figure 26. Histogram of Length of Initial Transient Period using the Schruben Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.95.

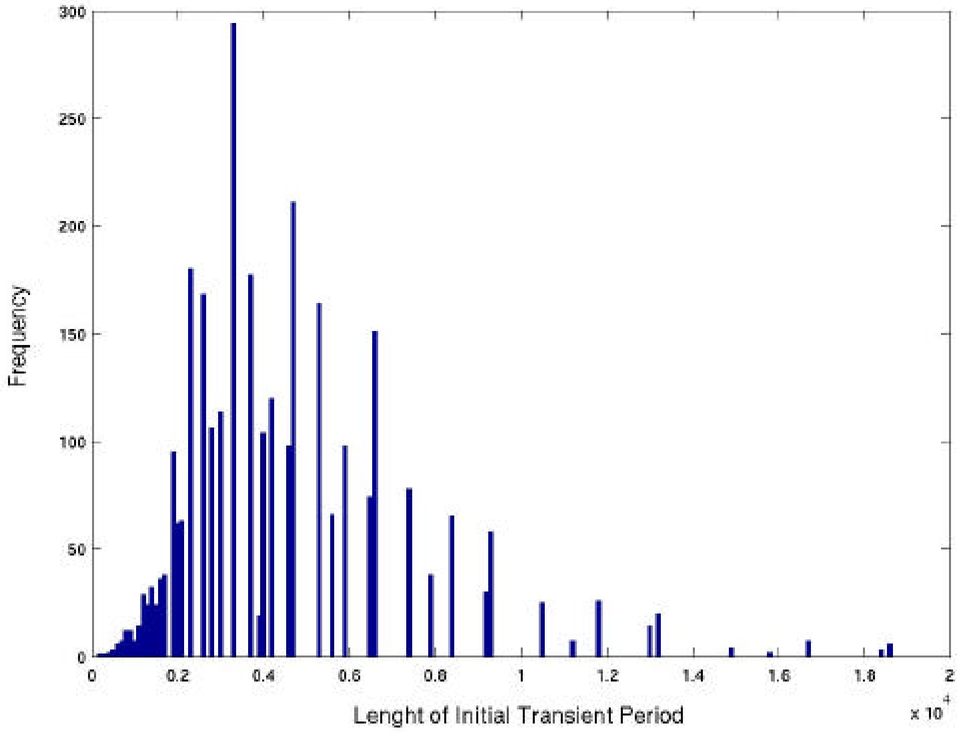

Figure 27. Histogram of Length of Initial Transient Period using the LWS Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.95.

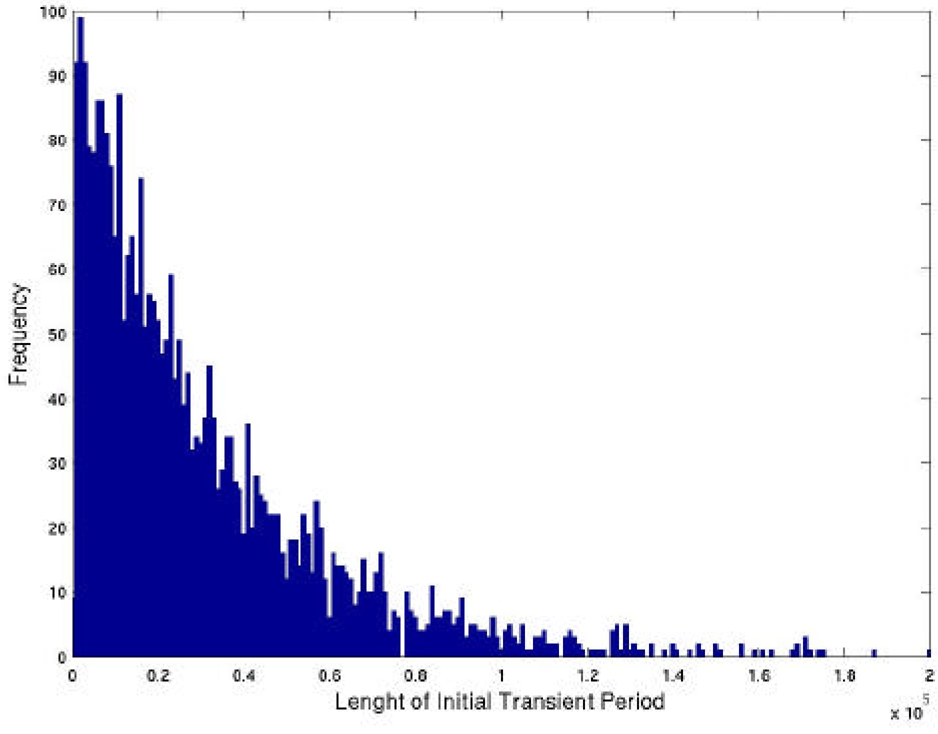

Figure 28. Histogram of Length of Initial Transient Period using the Rank Test with 3000 Runs for the M/M/1/8 Queue at ? = 0.95.

3.2.3. Evaluation of the GSS Tests

As detailed in Section 2.2.3, the GSS tests are based on an *F* statistic that compares the variability in the first portion of the output process (consist of b' batches) to the variability in the latter portion of the process (consist of b-b' of the remaining batches). We implemented the GSS tests using three different number of batches: 16, 8, and 4 (Cash et al [1992] recommended b = 16 as the maximum number of batches) and the sequential method described in Figure 13. The result of these experiments for the Area test, Maximum test, and Batch-mean test are shown in Figures 29, 30, and 31. These figures show that selecting 16 batches for all three tests leads to the largest estimate of the length of initial transient period. We therefore chose 16 in our future experiments as the number of batches used[10].

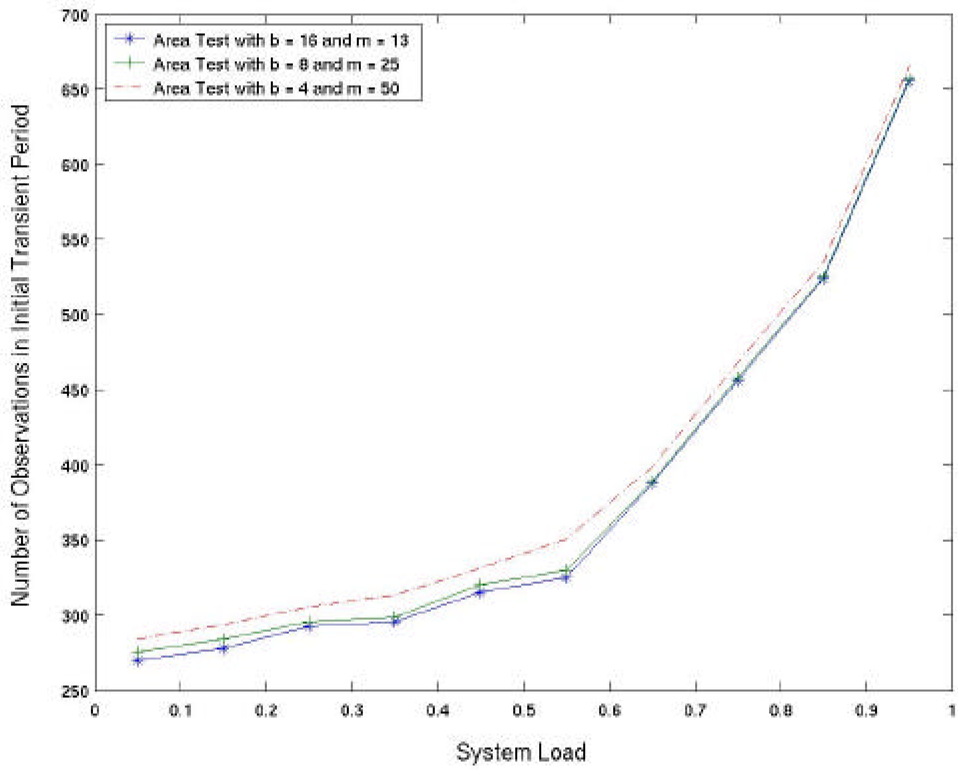

Figure 29. Evaluation of the Performance of the Area Test with Different Number of Batches in an M/M/1/8 Queueing System.

[10] It is more likely that the observations after the initial transient period represent steady state behaviour if we remove more observations from the beginning of each simulation run

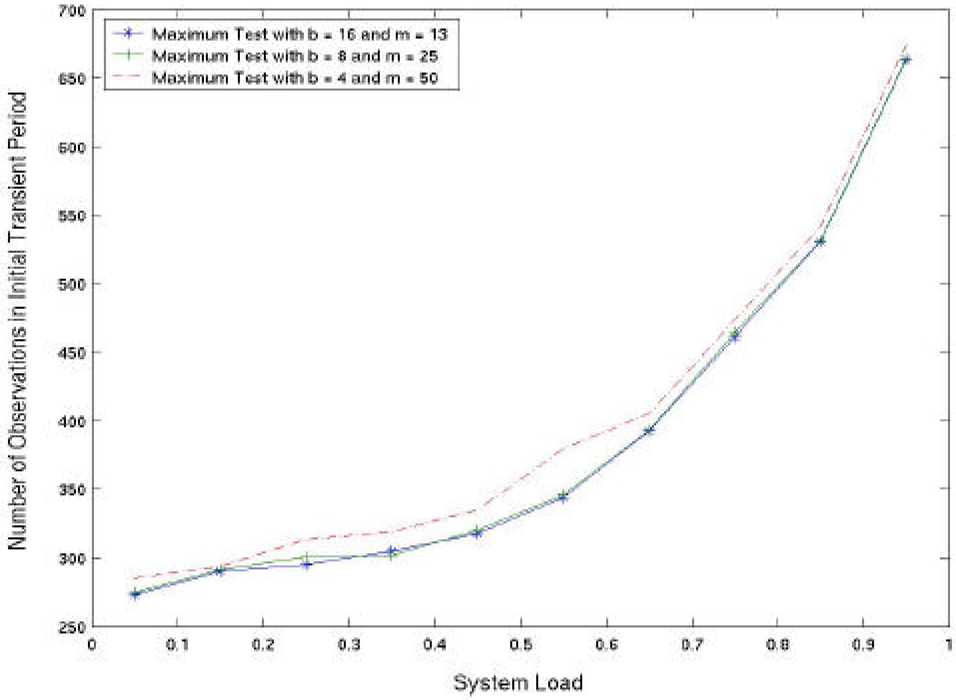

Figure 30. Evaluation of the Performance of the Maximum Test with Different Number of Batches in an M/M/1/8 Queueing System.

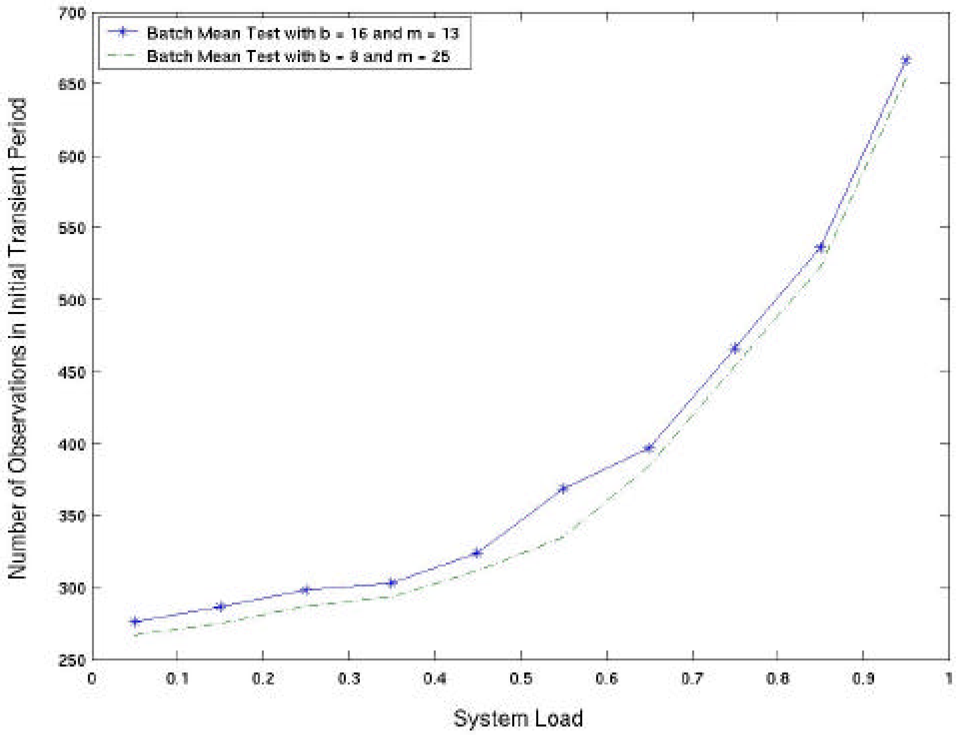

Figure 31. Evaluation of the Performance of the Batch-mean Test with Different Number of Batches in an M/M/1/8 Queueing System.

3.2.4. Evaluation of the Modified Schruben Test

We investigated four different modifications of the Schruben test, moving the variance window 1000, 10000, 20000, and 50000 observations away from the test window[11]. Figure 32 shows the result of these experiments. This figure shows that there is a slight difference between the length of the initial transient period detected by the Schruben test and the Modified Schruben test. In a heavier traffic load, the modified Schruben test detects a slightly longer length of initial transient period than that detected by the original Schruben test. The figure also shows that increasing the distance, *d*, of the variance window from the test window gives no obvious improvement.

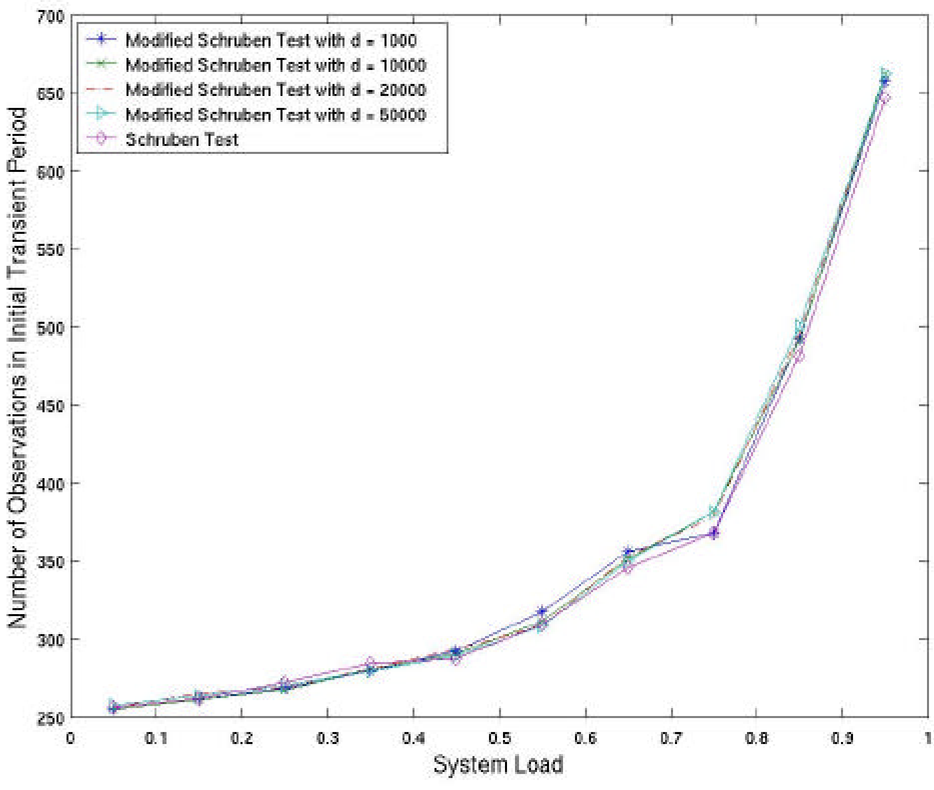

Figure 32. Comparison of the Performance of the Schruben Test with Four Different Versions of Schruben Test Modified for the M/M/1/8 Queue.

[11] We chose a large arbitrary number of 1000 as the minimum distance between the variance window and the test window so that there is more likely the variance is calculated over the observations in steady state.

3.2.5. Comparison of the Initial Transient Detectors with the Theoretical Estimates

In this section, we compare the length of the initial transient period in the mean waiting time of the customers in the system detected by the statistical tests and those produced from the Kelton and Law algorithm and the relaxation time in different system traffic loads. Table 9 shows the result of this comparison for the M/M/1/8 queue with no customer in the system at time zero. Comparing the length of initial transient period detected by two versions of Schruben test (with variable or fixed window size; see Section 2.2.1 and 2.2.2), the GSS tests, and the Rank test with the Kelton and Law and the relaxation time approximation, it can be seen that in the low traffic load, all the implemented statistical tests (except the LWS test) detect a much longer length of initial transient; greater than 200. This is because using the sequential method described in Figure 9, the minimum length of the initial transient period is equal to the window size, which is set to 200 here. This table also shows that the LWS test detects a much longer length of initial transient period than the other tests (except the Rank test) in a high traffic load. This is because the LWS test detects the length of initial transient period based on the independency of the batch means (i.e., in a high system load, the batch means are highly dependent).

Tests / Load	Schruben Test with Variable Window Size in Akaroa2	Schruben Test with Fixed Window Size	Batch-mean Test	Maximum Test	Area Test	Rank Test	LWS Test	Kelton and Law Algorithm	Relaxation Time
0.05	256.00	256.28	262.28	271.64	267.48	255.28	24.64	3.00	4.70
0.15	261.70	262.06	268.06	279.50	275.34	264.06	28.48	5.00	7.60
0.25	271.78	272.48	275.48	285.88	280.68	284.48	27.20	7.00	11.40
0.35	284.20	284.90	283.90	301.58	295.34	313.90	32.96	10.00	17.10
0.45	287.84	287.84	302.08	305.20	300.00	389.37	45.50	15.00	26.30
0.55	309.52	309.80	324.32	325.36	318.08	511.30	78.04	24.00	42.70
0.65	355.10	346.24	364.54	357.69	350.40	788.45	101.88	43.00	76.00
0.75	385.83	368.15	385.04	379.30	369.58	1668.33	160.64	90.00	159.10
0.85	542.82	481.51	467.98	471.35	461.82	4964.04	442.02	265.00	469.00
0.95	1016.27	647.16	586.37	591.38	583.44	30441.46	4522.58	2538.00	4456.40

Table 9. Length of the Initial Transient Period of the Mean Waiting Time of the Customers in the System Detected by the Statistical Tests and the Approximations Based on the Theory for Different System Load in the M/M/1/8 Queue.

Figures 33-41 show these comparisons more clearly for the M/M/1/8, M/Erlang$_4$/1/8 and M/Pareto$_{a=2.1}$/1/8 queues. As discussed in Chapter 2, underestimating the length of the initial transient period will leave bias in the steady state results so our aim here is to find the statistical test that gives the results close to those obtained theoretically by the relaxation time and the algorithm by Kelton and Law.

Figure 33 shows that this underestimation problem arises when the system is in the heavier traffic load over 0.85. This figure shows that the Schruben tests and the GSS tests underestimate the length of the initial transient period while the system is in heavy traffic load (for more detail see Figure 34). It can also be seen that the Rank test overestimates the length of the initial transient period by a very large margin from the results obtained by the relaxation time and the Kelton and Law algorithm, making this test very unreliable. For this reason, we did not include the Rank test in our further analysis of the M/Erlang$_4$/1/8 and M/Pareto$_{a=2.1}$/1/8 queueing systems.

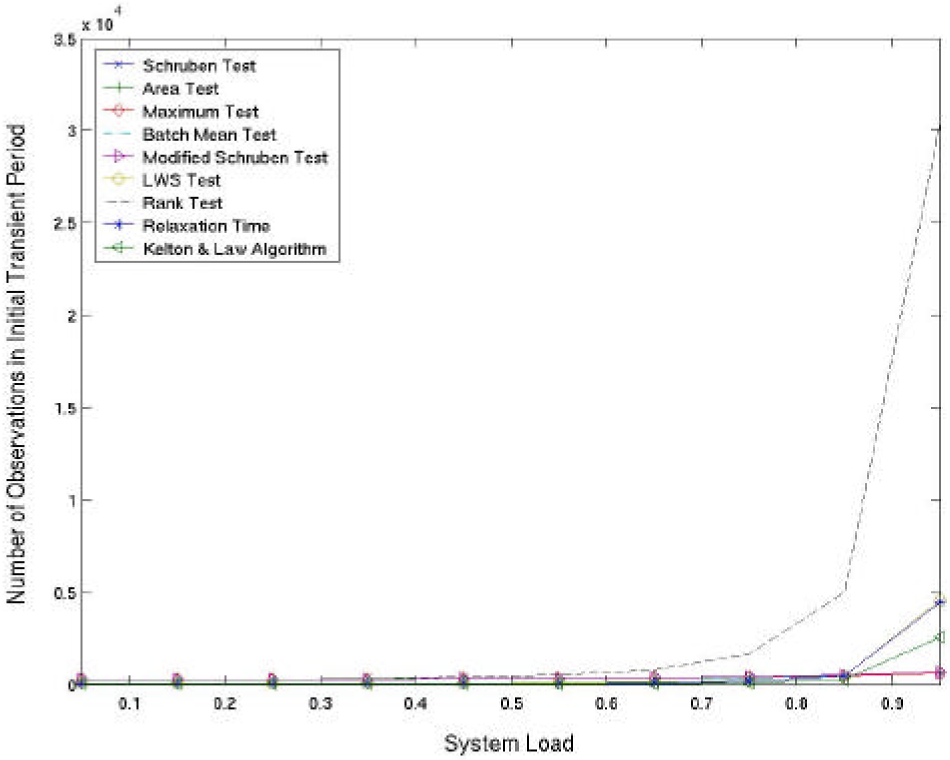

Figure 33. Comparison of the Statistical Tests with the Results Obtained by the Theory for the M/M/1/8 Queue.

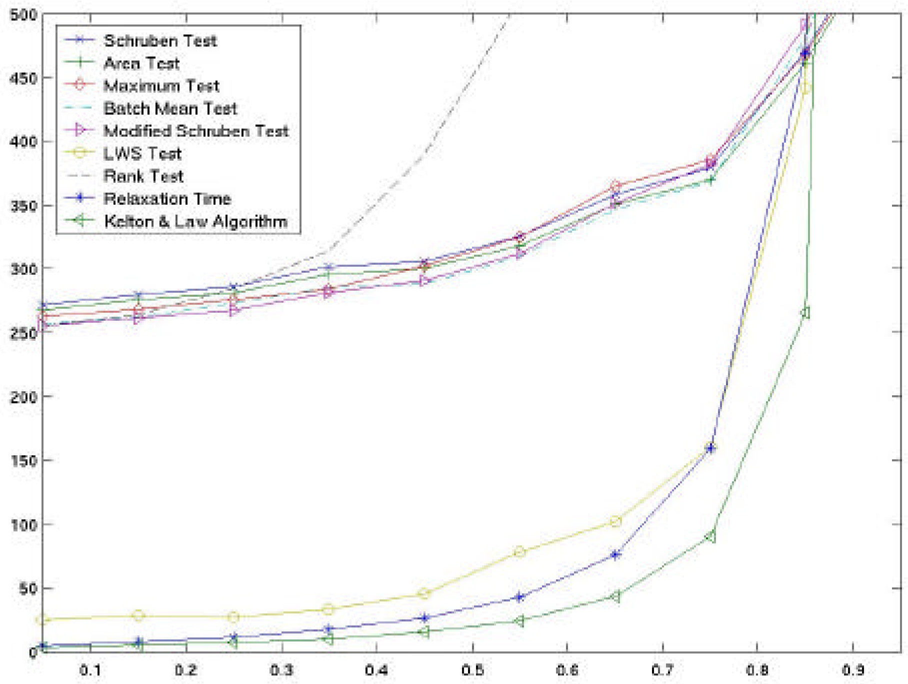

Figure 34. Comparison of the Statistical Tests with the Results Obtained by the Theory for the M/M/1/8 Queue.

Figure 35 takes a closer look at the performance of the Schruben tests and the GSS tests. This figure shows that the Schruben and the Modified Schruben tests detect a slightly longer length of the initial transient period for heavier traffic loads than the GSS tests.

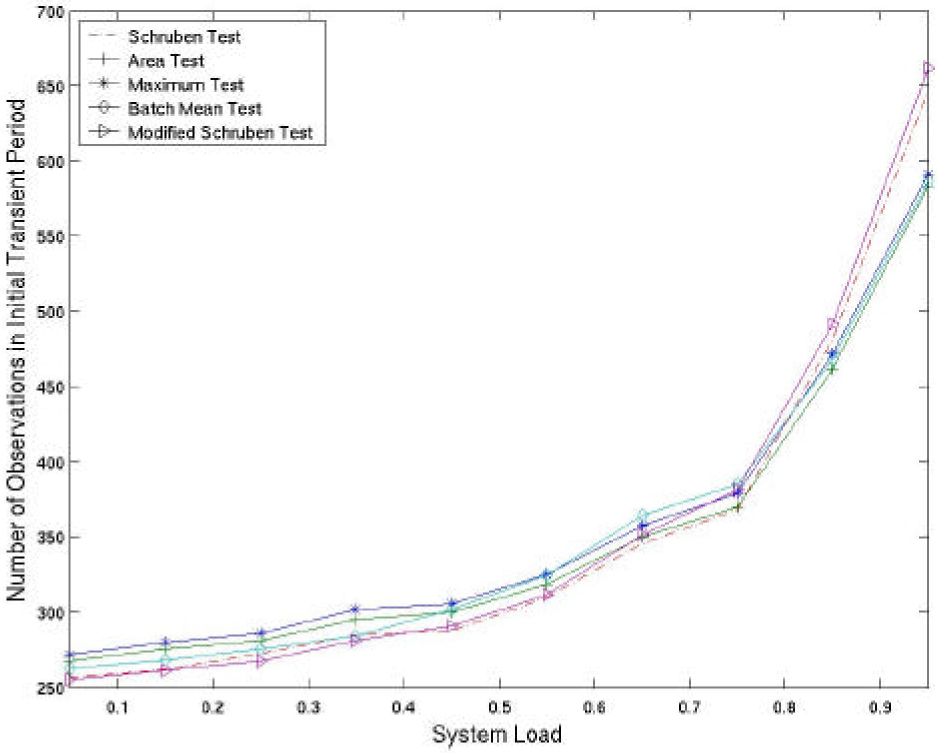

Figure 35. Comparison of the Results Obtained by the GSS Tests with the Schruben Test for the M/M/1/8 Queue.

Figure 36 compares the performance of the LWS test with the relaxation time and the results obtained by the Kelton and Law algorithm as also shown in Figure 33. It shows that the LWS test estimates the length of the initial transient period extremely close to that approximated by the relaxation time.

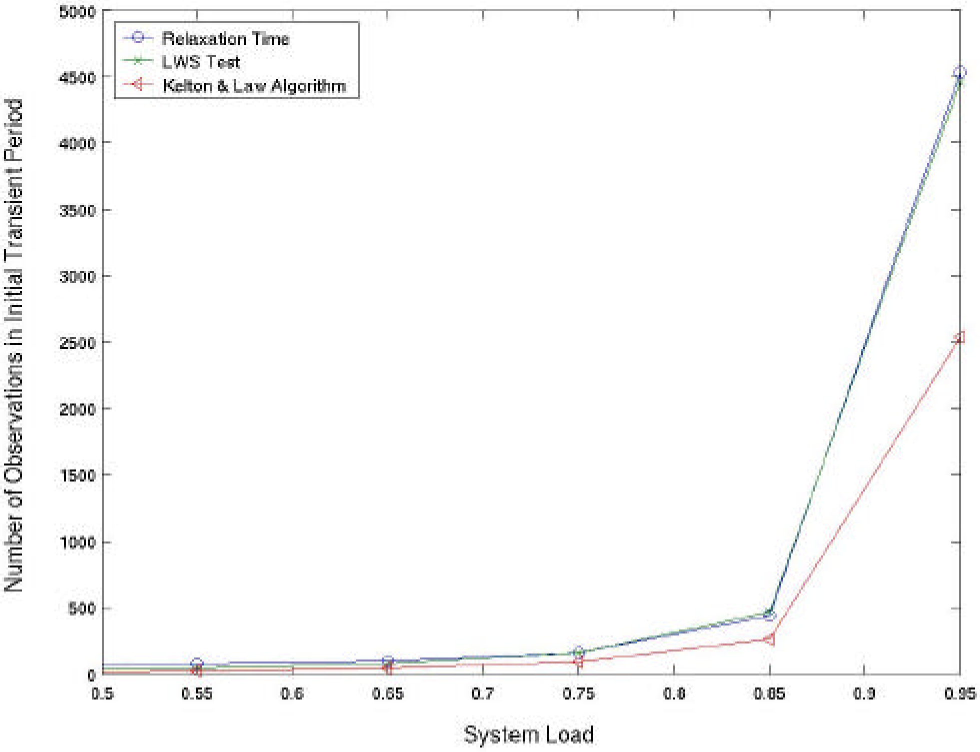

Figure 36. Comparison of the LWS Test with the relaxation Time and the Kelton and Law Algorithm for the M/M/1/8 Queue.

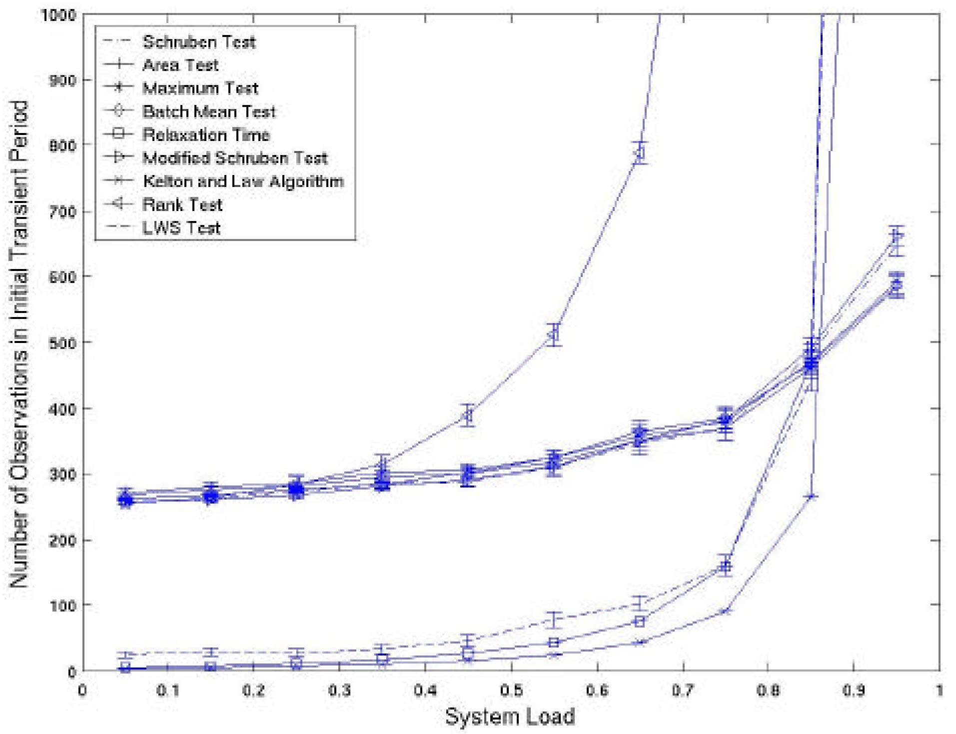

Figure 37. Performance of the Initial Transient Detectors in an M/M/1/8 Queue.

Figures 38-41 show the performance of these statistical tests (except the Rank test) and their comparison with the relaxation time in the case of M/Erlang$_4$/1/8 and M/Pareto$_{a=2.1}$/1/8 queueing systems.

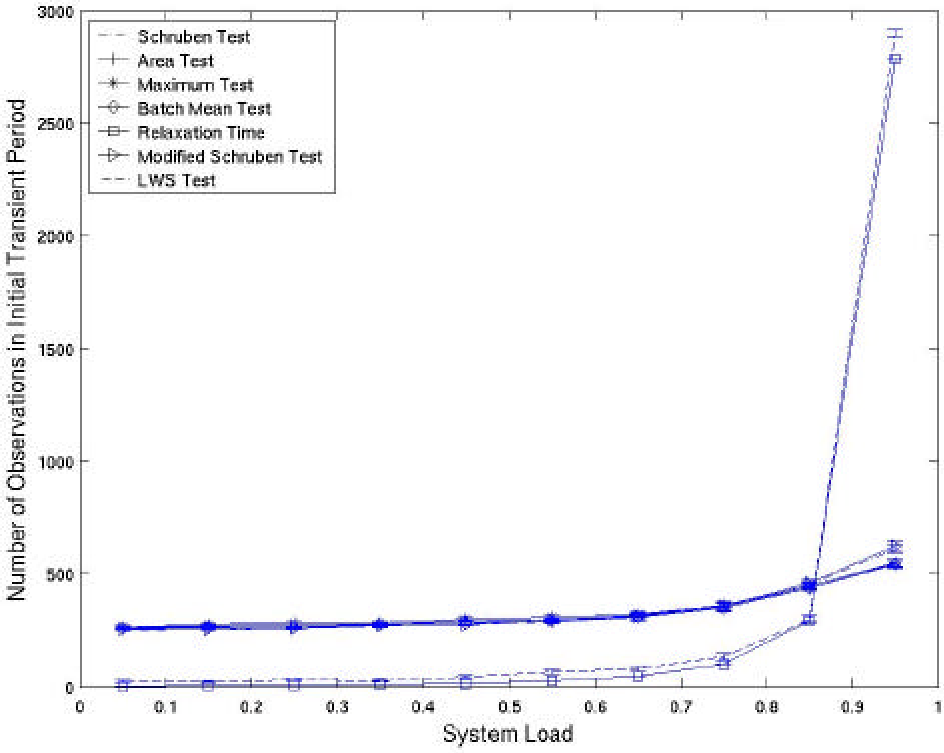

Figure 38. Performance of the Initial Transient Detectors in an M/Erlang$_4$/1/8 Queue.

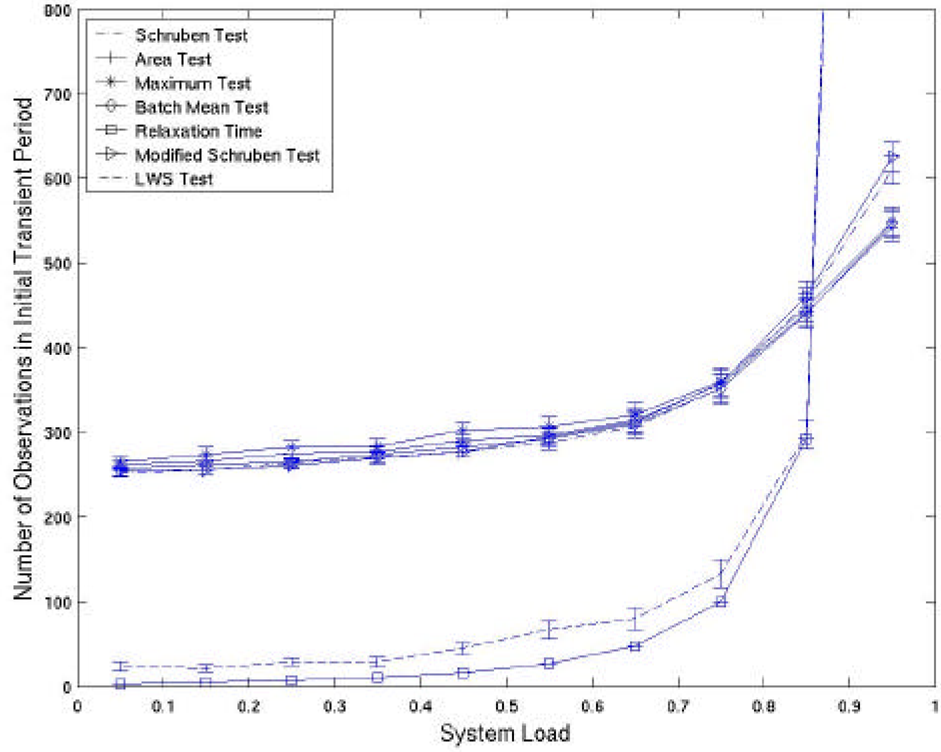

Figure 39. Performance of the Initial Transient Detectors in an M/Erlang$_4$/1/8 Queue.

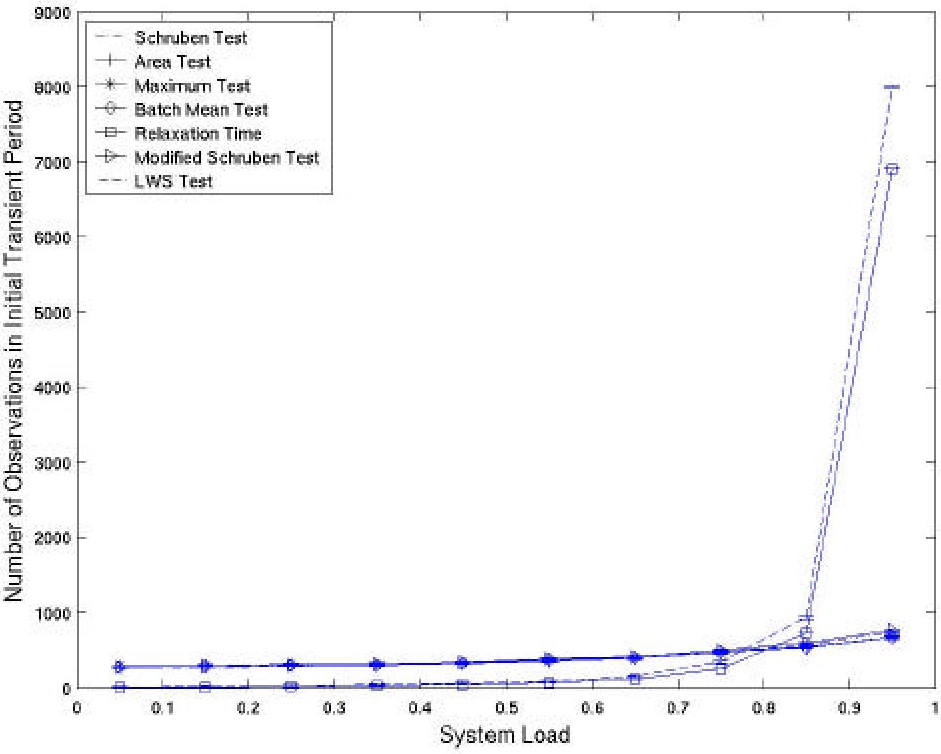

Figure 40. Performance of the Initial Transient Detectors in an M/Pareto$_{a=2.1}$/1/8 Queue.

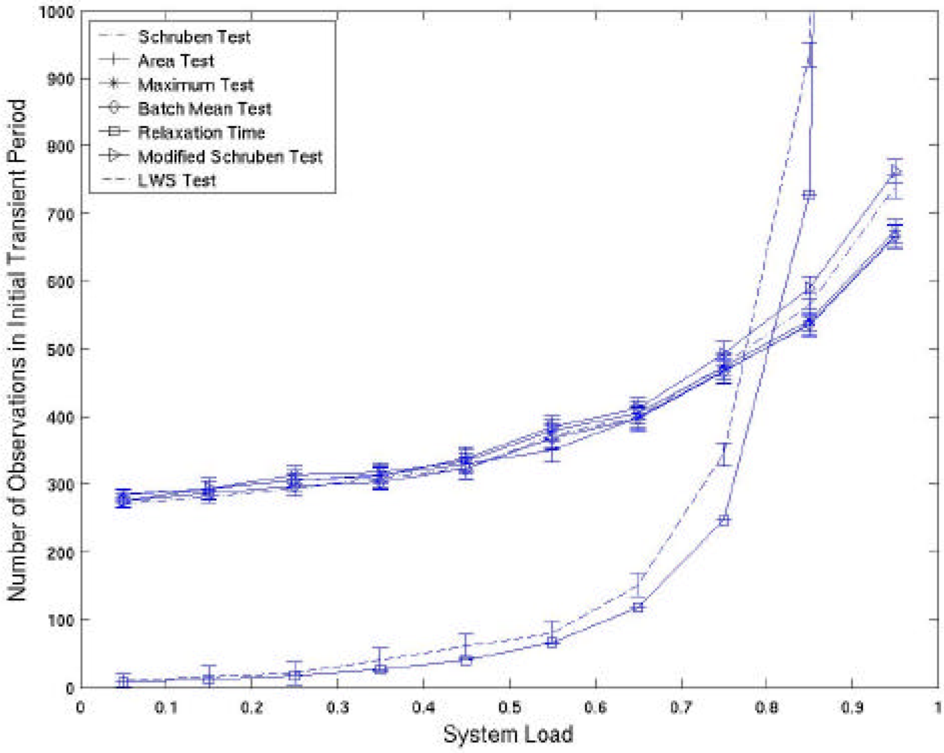

Figure 41. Performance of the Initial Transient Detectors in an M/Pareto$_{a=2.1}$ /1/8 Queue.

3.2.6. Investigation on the Accuracy of the Statistical Tests

So far, we have looked at the comparison of the tests in terms of the length of initial transient period they detect. In this section, we investigate the accuracy of these detections by checking if the system can be said to be in steady state or not after the estimated transient period. To do this, we compare the mean waiting time of the first 500 customers[12] immediately after the end of the initial transient period with the steady state waiting time calculated by Pollaczek-Khintchine formula,

$$w = \frac{1}{m} + \frac{1}{l}\left(\frac{r^2}{1-r}\right)\left(\frac{1+C_S^2}{2}\right),$$

where m is the service rate (equal to 10), l is the arrival rate, r is the system load, and C_S^2 is the coefficient of variation for the service time. The comparison of these average waiting times with the steady state waiting times calculated from the Pollaczek-Khintchine formula are presented in Table 10.

[12] This analysis was initially performed as our first experiment with an arbitrary number of 500 customers immediately after the end of the initial transient period. However, it was then improved by the comparison of the CDF of the first customer immediately after the end of the initial transient with the steady state cumulative distribution function (discussed in further detail later in this section).

Test / Load	Schruben Test	Batch-mean Test	Area Test	Maximum Test	LWS Test	Rank Test	Steady state Waiting Time
0.0500	0.1068	0.1063	0.1050	0.1052	0.1053	0.1062	0.1052
0.1500	0.1195	0.1179	0.1180	0.1178	0.1193	0.1183	0.1176
0.2500	0.1348	0.1318	0.1299	0.1321	0.1327	0.1328	0.1333
0.3500	0.1506	0.1511	0.1533	0.1501	0.1551	0.1507	0.1538
0.4500	0.1797	0.1729	0.1748	0.1769	0.1808	0.1798	0.1818
0.5500	0.2256	0.2184	0.2233	0.2219	0.2191	0.2166	0.2222
0.6500	0.2895	0.2924	0.2882	0.2842	0.2839	0.2588	0.2857
0.7500	0.3855	0.4179	0.4126	0.4070	0.3948	0.3465	0.4000
0.8500	0.5944	0.6850	0.6679	0.6497	0.7026	0.4510	0.6666
0.9500	1.4932	1.6195	1.5756	1.5910	2.0503	0.7554	2.0000

Table 10. Mean Waiting Time of 500 Customers after Initial Transient Period is Finished Averaged of 1000 Replications for the M/M/1/8 Queue.

Table 10 shows that, with a low traffic load, the value of the mean waiting time of the first 500 customers after the end of the initial transient period is very close, but often larger than the actual steady state value. In contrast, in a higher traffic load, for all the tests (except the LWS test) this mean waiting time becomes smaller than the steady state value. Looking at these results for the Rank test, the mean waiting time in a high traffic load is much smaller than the steady state value (almost half of the steady state value). To clarify this behaviour of the Rank test, we drew 10 different simulated waiting times of the *n*th customer in an M/M/1/8 queue. Figures 42-45 show four of these simulated waiting times. As shown in these figures, the Rank test generally detects the length of the initial transient period when the queue is empty or close to empty, and when the waiting time is very small. This explains the results that were presented in Table 10 and gives one more reason why the Rank test is not to be recommended.

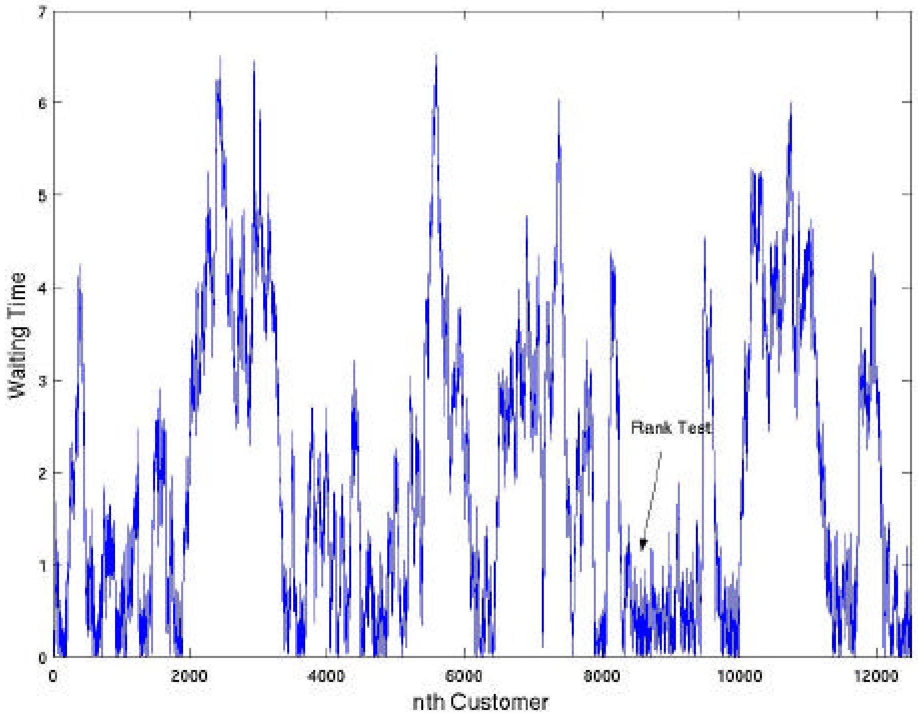

Figure 42. The Length of the Initial Transient Period equal to 8696 Detected by the Rank Test for an M/M/1/8 Queue with ? = 0.95.

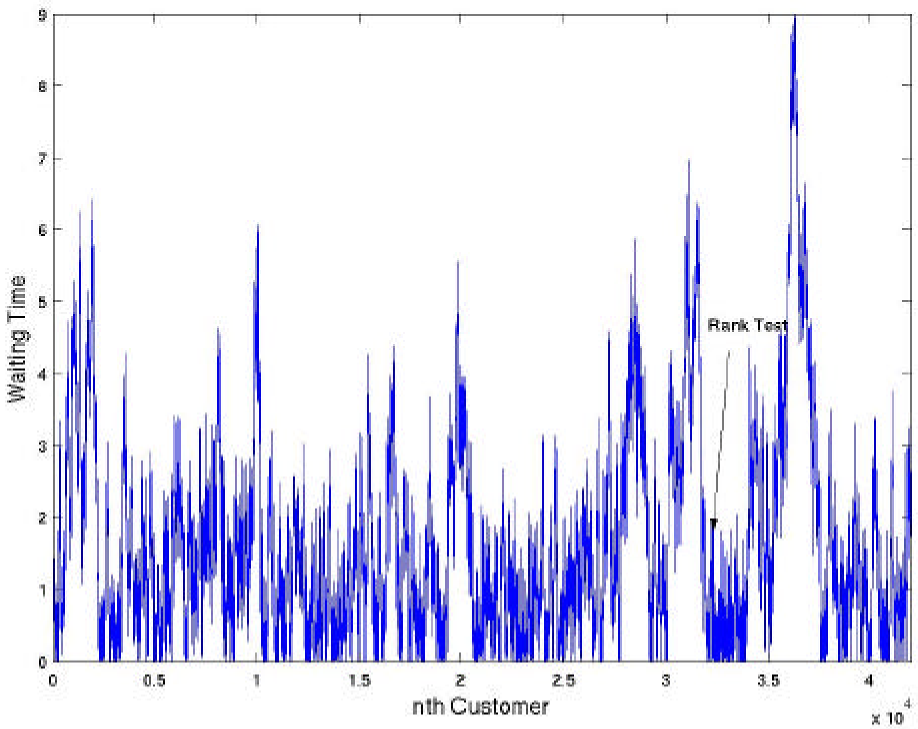

Figure 43. The Length of the Initial Transient Period equal to 32653 Detected by the Rank Test for an M/M/1/8 Queue with ? = 0.95.

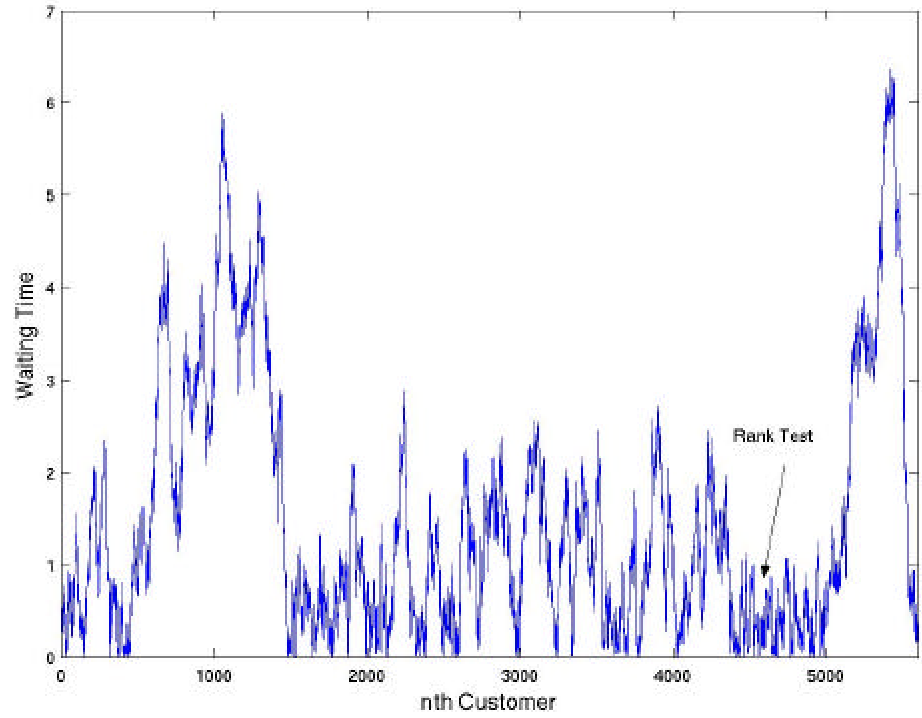

Figure 44. The Length of the Initial Transient Period equal to 4627 Detected by the Rank Test for an M/M/1/8 Queue with ? = 0.95.

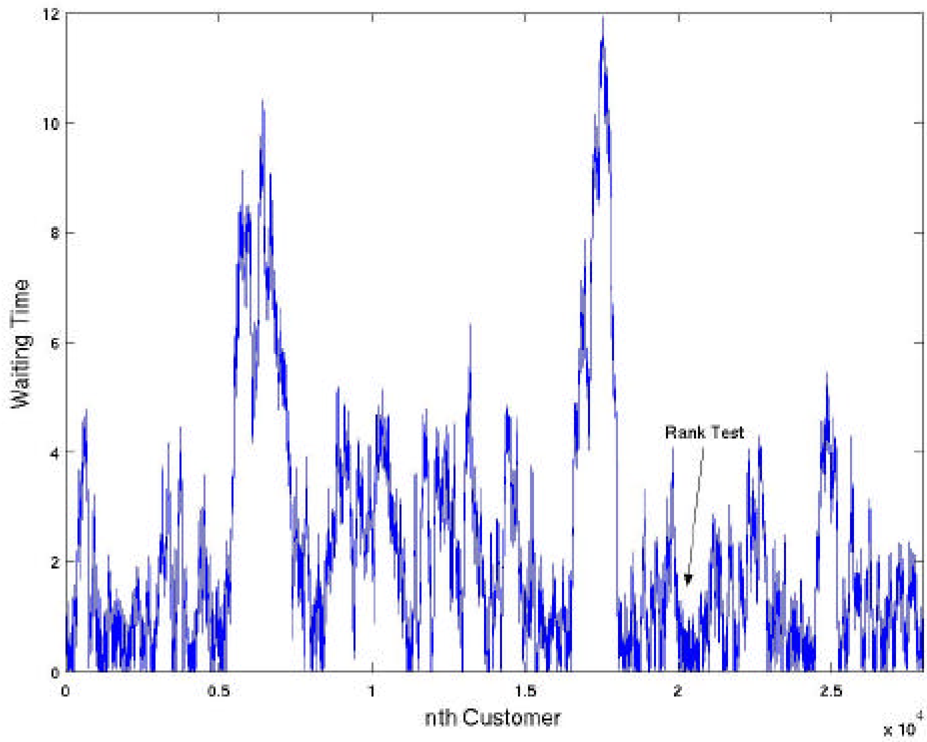

Figure 45. The Length of the Initial Transient Period equal to 20380 Detected by the Rank Test for an M/M/1/8 Queue with ? = 0.95.

As a further measure, we calculated the waiting time of the first customer, supposedly in steady state, after discarding the initial transient observations according to the Area test, the Maximum test, the Batch-mean test, the Rank test, the Schruben test implemented in Akaroa2 with a variable window size, the Schruben test with a fixed window size, and the LWS test. The results were observed from 1000 runs for three different queueing systems: M/M/1/8, M/Erlang$_4$/1/8, and M/Pareto$_{\alpha=2.1}$/1/8 with the service rate equal to 10. The mean of these waiting times are presented in Table 11.

System / Test	**M/M/1/8**	**M/Erlang$_4$/1/8**	**M/Pareto$_{a=2.1}$/1/8**
Area Test	0.9398	0.6405	2.4727
Maximum Test	0.9575	0.6441	2.4539
Batch-mean test	0.9547	0.6691	2.4886
Schruben Test	0.8276	0.5497	2.2219
Schruben Fixed Test	0.7846	0.5638	2.2266
LWS Test	0.9601	0.6429	2.4628
Rank Test	0.5261	0.4185	1.4476
Steady state Waiting Time	1.0000	0.6625	2.6928

Table 11. The Mean Waiting Time of the first Customer in Steady state Period, Averaged over 1000 Runs for System Load 0.9.

Table 11 shows that the mean waiting time in the system of the first customer, supposedly in steady state, after discarding the initial transient observations according to the Rank test, is significantly smaller than the other tests. This agrees with the results presented in Figures 43-46; the Rank test tends to detect the end of the initial transient period when the queue is empty or close to empty. We compared these results to the steady state mean waiting time calculated from the Pollaczek-Khintchine formula, with C_S^2 the coefficient of variation of the service time equal to 1, $1/k$ and $1/a(a-2)$ for the M/M/1/8, M/Erlang_k/1/8 and $\text{M/Pareto}_{\alpha=2.1}$/1/8 queueing systems, respectively. Setting the mean service time to 1/9, and r = 0.9, the mean steady state waiting time for the M/M/1/8, M/Erlang_4/1/8 and $\text{M/Pareto}_{\alpha=2.1}$/1/8 systems are 1, 0.66 and 2.69, respectively. Comparing these mean steady state waiting times with the observed results in Table 11, we can state that the mean waiting time of the first customer in steady state detected by the GSS tests and the LWS test are the closest to the steady state waiting times for all the three queueing systems. This table also shows that the results for the Rank test are the furthest from the steady state waiting time, almost half of these theoretical values. We also show that the Schruben test with the variable window size outperforms the Schruben test with the fixed window size only in the case of the M/M/1/8 queueing system.

We recorded the waiting time of the next customer after the end of the transient period for 1000 independent customers (from 1000 independent runs each starting with a different seed) for a M/M/1/8 queue with a system load of 0.9 using the above tests. In addition to the statistical tests, we performed this experiment on the same 1000 independent runs by removing a fixed number of observations calculated by the relaxation time (1084 observations) and the algorithm by Kelton and Law (615 observations) from the beginning of each run. Ordering these 1000 independent waiting times in ascending order, we estimated the cumulative distribution function of these waiting times in the system and compared them to the steady state cumulative distribution function (CDF) for an M/M/1/8 queue given as $1-e^{-m(1-r)t}$.

The results are presented in Figures 46-52. The observed and the theoretical steady state CDFs are compared using the Kolmogorov-Smirnov Test. Using the Kolmogorov-Smirnov Test, one can determine whether two data sets are drawn from the same distribution. Given the hypothesized continuous distribution function *F*, this test compares *F* to the empirical distribution function, *F¢*, of the samples. The Kolmogorov-Smirnov test statistic *D* is the largest absolute deviation between *F(x)* and *F¢(x)* over the range of the random variable:

$$D = \max_x \{| F'(x) - F(x) |\}$$

A line labelled as KS shows this largest absolute deviation in Figures 46-54. As Figures 46 and 47 show, the CDFs of the relaxation time and the algorithm by Kelton and Law are extremely close to the theoretical steady state CDF. This shows that deleting a fixed number of observations obtained by the theory gives accurate results. However, as was mentioned in Chapter 2, the theoretical results can not be obtained for all the queueing models.

We show in the remaining figures that the CDF of the waiting times, while using the Rank test to detect the length of initial transient, is the furthest from the steady state CDF. This is contrary to the results by the GSS and LWS tests that are very close to the steady state CDF.

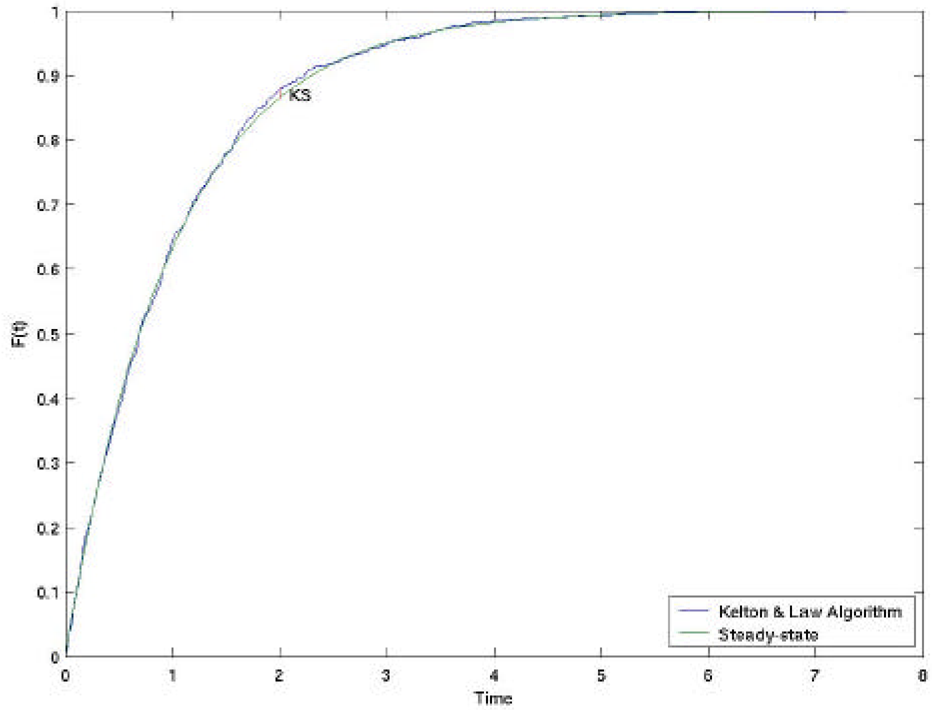

Figure 46. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the Fixed Theoretical Value Calculated by the Kelton and Law Algorithm to Detect the Length of the Initial Transient Period.

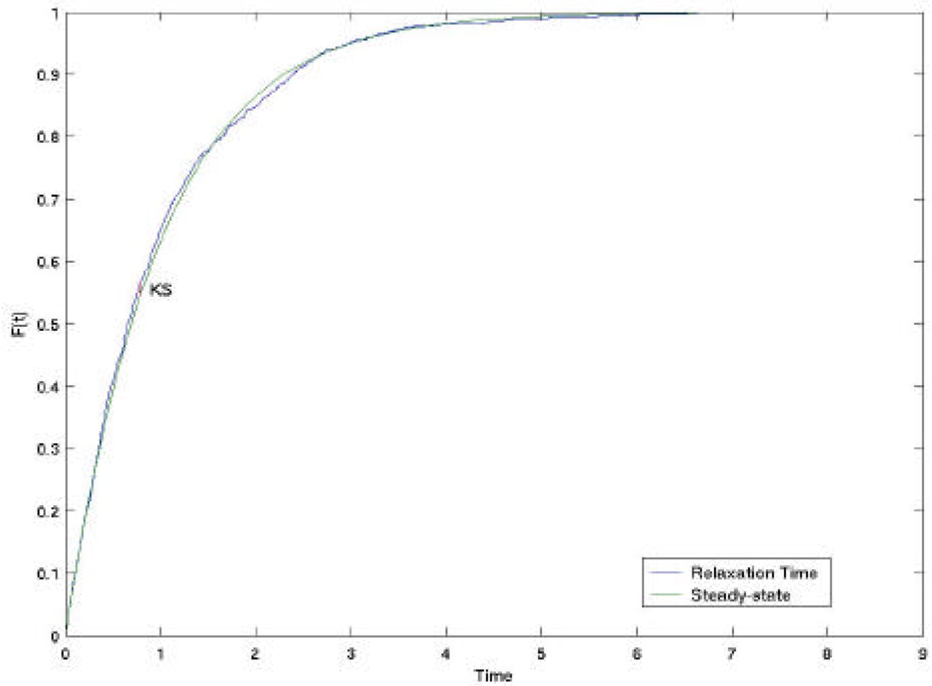

Figure 47. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the Fixed Theoretical Value Calculated by the relaxation Time to Detect the Length of the Initial Transient Period.

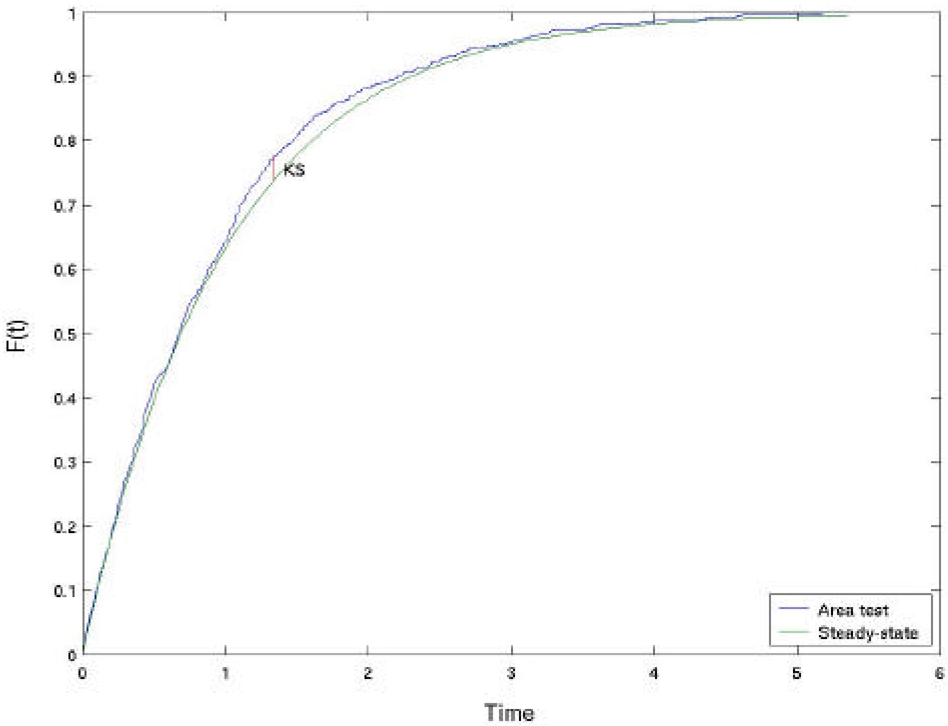

Figure 48. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the Area Test to Detect the Length of the Initial Transient Period.

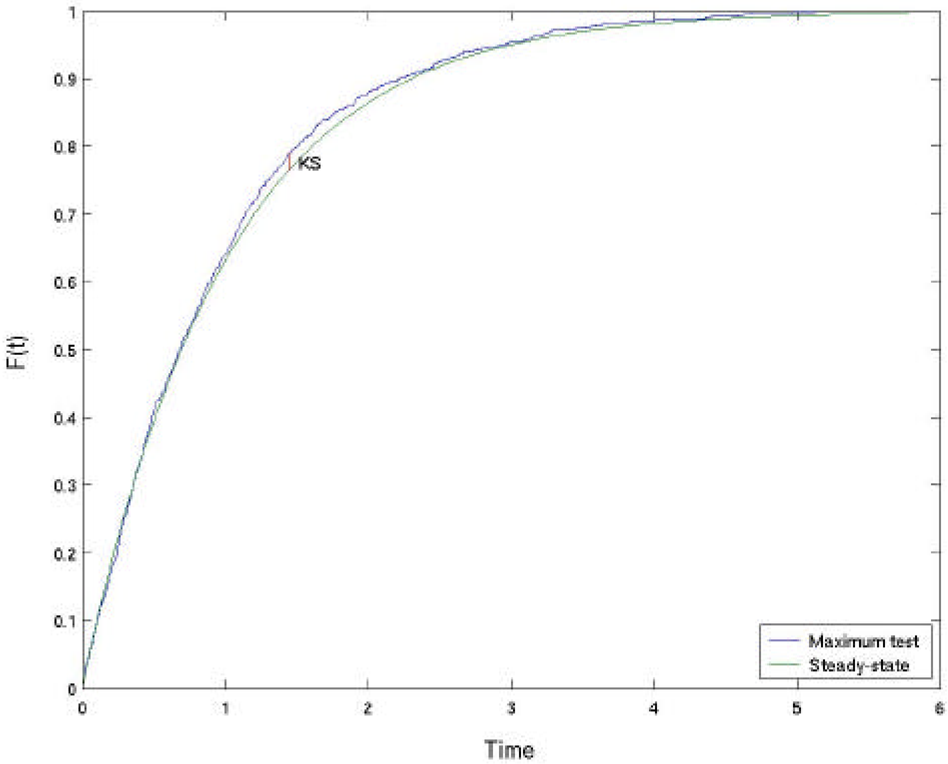

Figure 49. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the Maximum Test to Detect the Length of the Initial Transient Period.

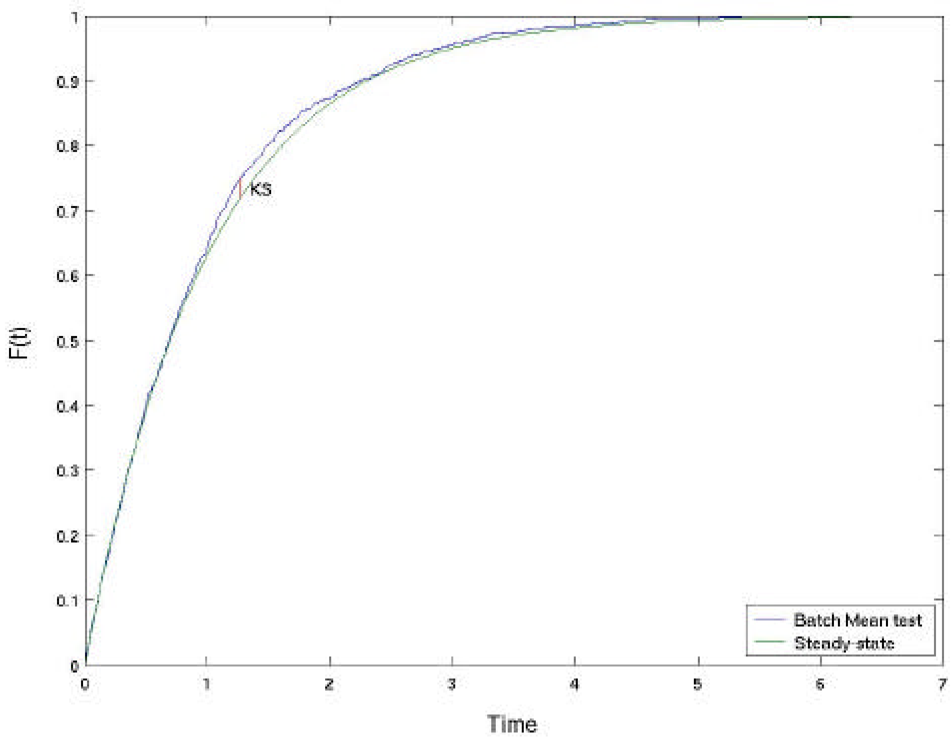

Figure 50. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the Batch-mean test to Detect the Length of the Initial Transient Period.

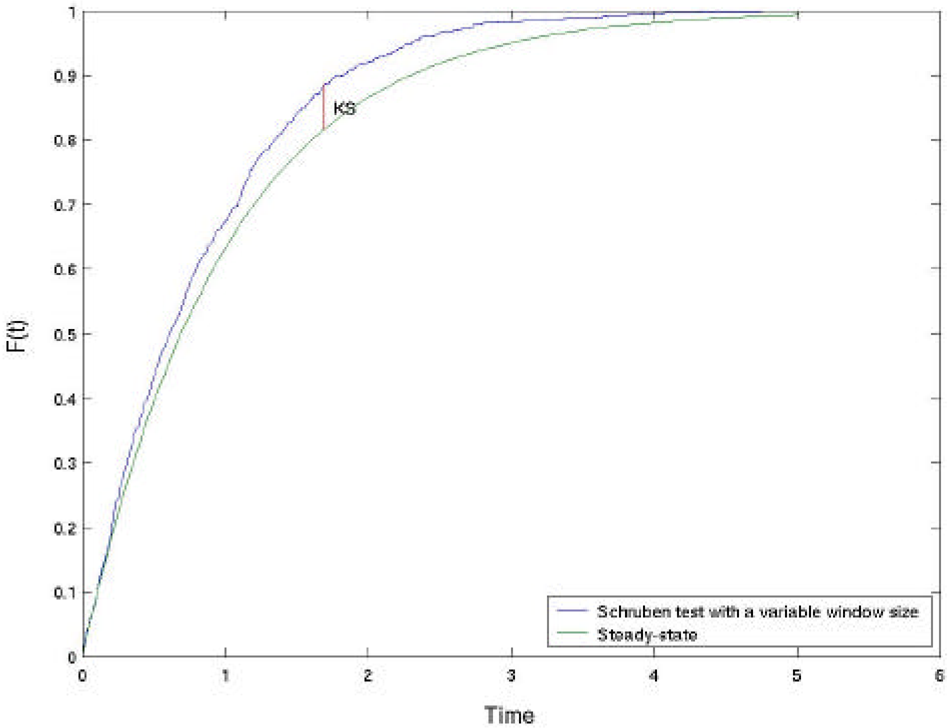

Figure 51. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the Schruben Test with Variable Window Size to Detect the Length of the Initial Transient Period.

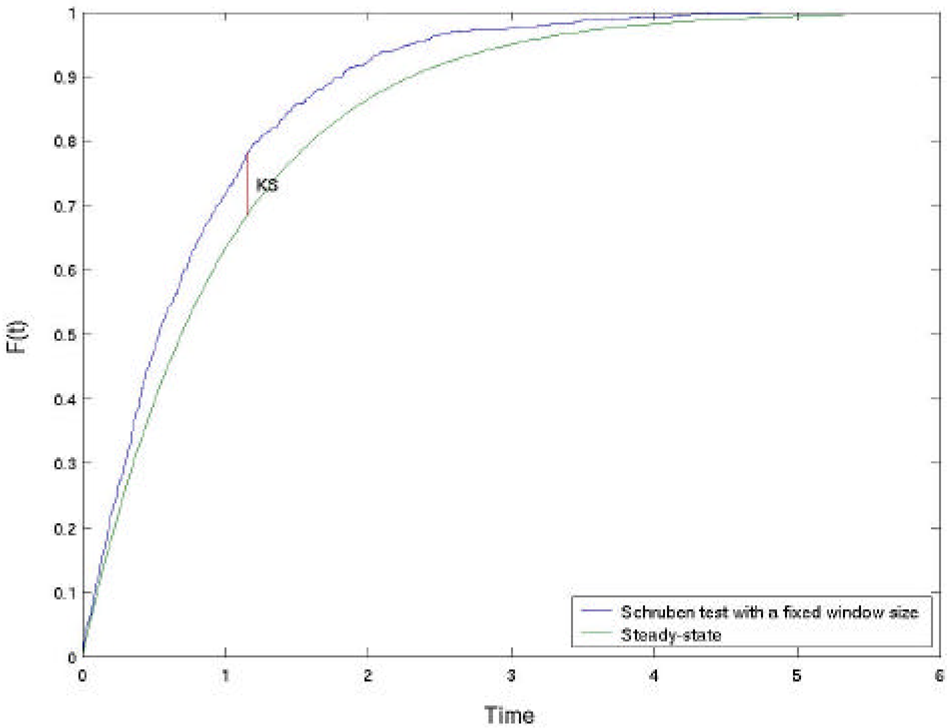

Figure 52. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the Schruben Test with Fixed Window Size to Detect the Length of the Initial Transient Period.

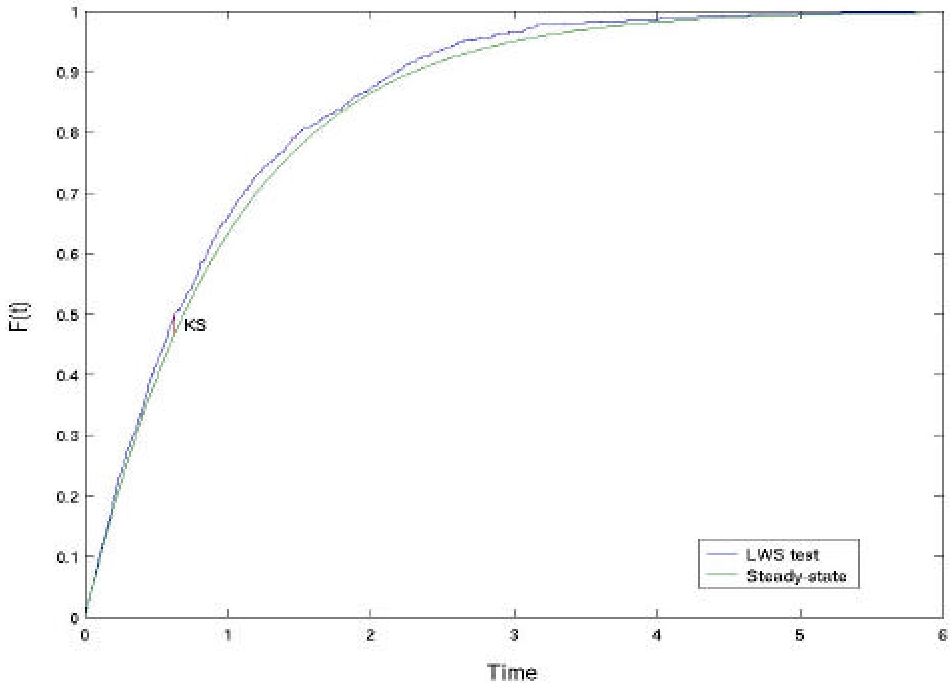

Figure 53. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the LWS Test with Fixed Window Size to Detect the Length of the Initial Transient Period.

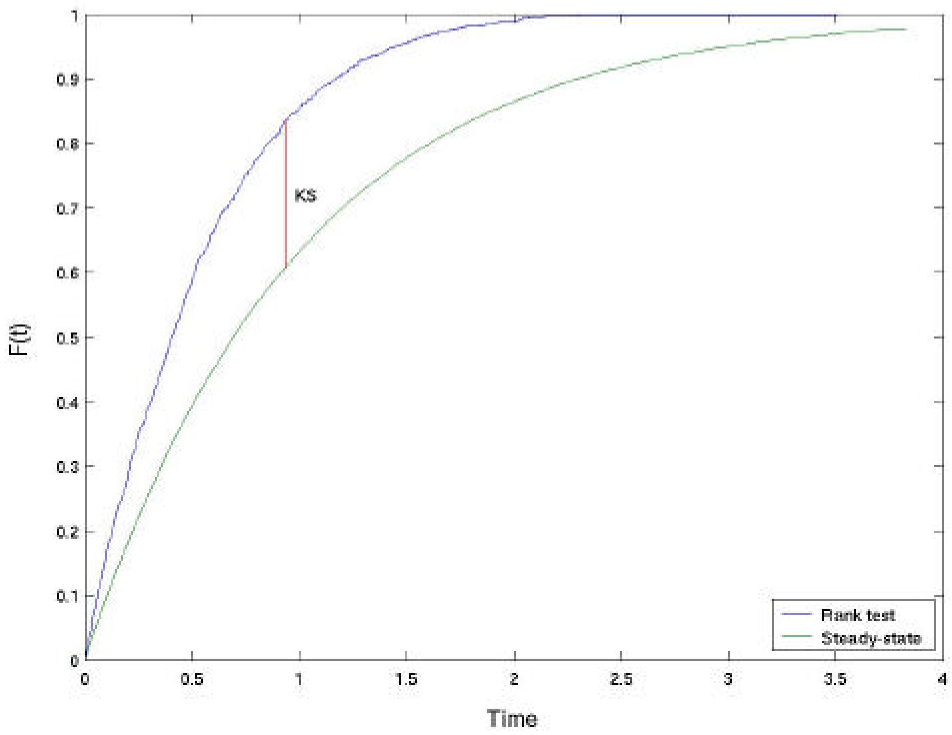

Figure 54. Comparing the Observed and M/M/1/8 Waiting Time in System CDFs using the Rank Test to Detect the Length of the Initial Transient Period.

3.2.7. Artificially Generated Datasets

So far we have investigated the performance of the statistical tests on three simple queueing models: M/M/1/8, M/Erlang$_k$/1/8 and M/Pareto$_\alpha$/1/8. In this section, we will test the performance of the Schruben, the LWS, and the Area tests (which have had a satisfactory performance in the previous experiments) on the six artificially generated stochastic processes as used by Bause and Eickhoff [2003] (see Figure 55). These are selected to investigate different type of convergence to steady state, or, as in the case of (*e*) and (*f*), no convergence to steady state at all. This is done by introducing bias to l initial number of observations of a set of normal random processes, using one of the transient functions (T_t) defined in Sections 3.2.7.1 – 3.2.7.6. By doing this, the length of the initial transient period is known in advance and so by running the tests on these processes, we can determine if the tests detected a correct length of the initial transient period.

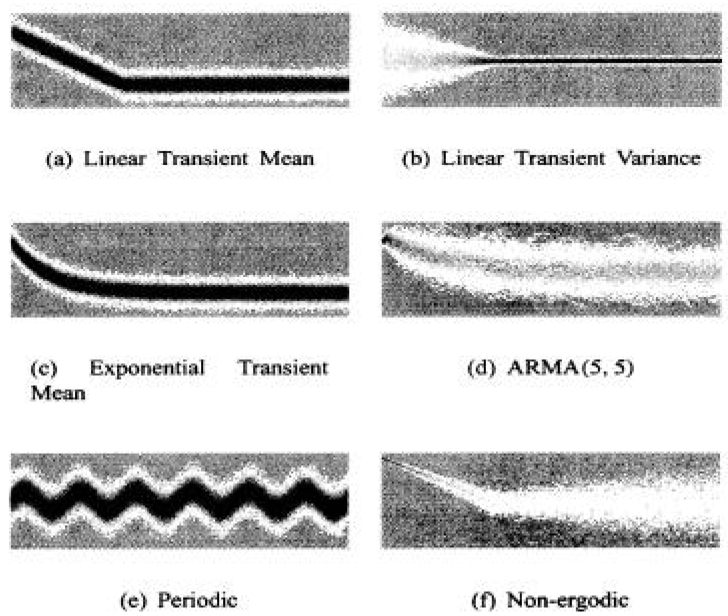

Figure 55. The Artificial Generated Stochastic Sequences used by Bause and Eickhoff [2003].

For all of the processes described below, let N_t be the realisation of an independent normal random process with mean 0 and variance 1. The number of independent replications was decided by using the sequential method described in Section 3.2.1 as was done previously. Here we chose 1.0 as our absolute error so that enough replications are collected when the length of initial transient period of the artificially generated processes is out by ±1 observation. We chose 100 and 10 as the values for the window size and the step length (for the Schruben test), and different lengths of initial transient period; smaller, equal, and larger than the window size (see Table 12). Following are the six well-known processes described and used by Bause and Eickhoff [2003].

3.2.7.1. Linear Transient Mean (LTM)

The first dataset, which can be seen in Figure 55(a), is a process with the linear transient mean where transient effects diminish to zero linearly as time approaches $t = l$. As described in Bause and Eickhoff [2003], the linear transient mean dataset is a realisation of the process

$$Y_t^{LTM} = T_t^{LTM} + N_t \,,$$

where T_t^{LTM} is defined as

$$T_t^{LTM} = \begin{cases} x - t\dfrac{x}{l} & if\ t < l \\ 0 & else \end{cases}$$

The role of x here is to set the amount of the difference between two consecutive random samples, so for example with a smaller value of x, the two consecutive random samples differ by a smaller value. We chose 10 for the value of x here (as used by Bause and Eickhoff [2003]) and considered different values for l, which is the point that prior to this point, there was bias introduced to the random sample (see Table 12).

3.2.7.2. Linear Transient Variance (LTV)

The Linear transient variance dataset, as shown in Figure 55(b), is a realisation of the process

$$Y_t^{LTV} = T_t^{LTV} \times N_t ,$$

where T_t^{LTV} is the transient process defined by the linear function

$$T_t^{LTV} = \begin{cases} x - t\dfrac{x-1}{l} & \text{if } t < l \\ 1 & \text{else} \end{cases}$$

The linear transient variance process has a constant mean, but there is a transient behaviour of the variance. We chose 10 for the value of *x* and considered different values for *l*, the length of the initial transient period (see Table 12).

3.2.7.3. Exponential Transient Mean (ETM)

The exponential transient mean dataset (see Figure 55(c)) where the initial transient period disappears exponentially is a realisation of the process

$$Y_t^{ETM} = T_t^{ETM} + N_t ,$$

where T_t^{ETM} is defined as

$$T_t^{ETM} = x \cdot e^{(t\frac{\ln(0.05)}{l})}$$

As it is stated in Bause and Eickhoff [2003], this transient function results in permanent differences between two consecutive random samples. However, beyond time index l, a test sample will differ from the steady state distribution at most by 5% (as chosen by Bause and Eickhoff [2003]) of the difference between the first random sample, Y_0^{ETM}, and the steady state distribution. Therefore, there is no clear truncation point in the case of the ETM processes. Here again, the value for x was chosen as 10 and we considered different values for l, the length of initial transient period (shown in Table 12).

3.2.7.4. ARMA(5,5)

The ARMA(p,q) dataset (see Figure 55(d)) includes both autoregressive (AR(p)) and moving average (MA(q)) terms, and, as described in Hamilton [1994], is a realisation of the process

$$Y_t^{ARMA(p,q)} = c + f_1 Y_{t-1} + f_2 Y_{t-2} + \ldots + f_p Y_{t-p} + N_t + q_1 N_{t-1} + q_2 N_{t-2} + \ldots + q_q N_{t-q}\,,$$

where $f_1, f_2, \ldots, f_p$ are the autoregressive parameters and $q_1, q_2, \ldots, q_p$ are the moving average parameters.

Here, we selected $p = q = 5$, $c = 1$, $f_i = q_i = \frac{1}{2^i}$, $t \geq 0$ and $Y_{-5}^{ARMA(5,5)} = Y_{-4}^{ARMA(5,5)} = Y_{-3}^{ARMA(5,5)} = Y_{-2}^{ARMA(5,5)} = 0, \quad Y_{-1}^{ARMA(5,5)} = 100$.

and so

$$Y_t^{ARMA(5,5)} = 1 + N_t + \sum_{i=1}^{5} \frac{1}{2^i}(Y_{t-1}^{ARMA(5,5)} + N_{t-i}), \quad t \geq 0$$

As stated in Bause and Eickhoff [2003], theoretically the sample mean at $l = 183$ differs at most by 5% from the steady state distribution.

3.2.7.5. Periodic

The periodic process is non-stationary, so it does not converge to steady state. The tests should therefore not detect a truncation point. This is a realisation of the process

$$Y_t^P = P_t^P + N_t \,,$$

where P_t^P is the periodic process defined as

$$P_t^P = b \cdot \sin(w\, t)$$

where $w = \dfrac{2p}{T}$ and T is the cycle length (see Figure 55(e)). We chose 103, a prime value larger than the window size, as the value of the cycle length.

3.2.7.6. Non-Ergodic

A non-ergodic process is one that if the process goes into an unusual state in one period or other, it will never return to normal or usual state (see Figure 55(f)). The non-ergodic process, as described in Bause and Eickhoff [2003], is a realisation of the process

$$Y_t^{NE} = NE_t \times N_t + T_t^{LTM} \,,$$

where $NE_t = 0.01t$, and $T_t^{LTM} = \begin{cases} x - t\dfrac{x}{l} & if\ t < l \\ 0 & else \end{cases}$

Again, we considered x = 10 and different values for l as can be seen in Table 12. This non-ergodic process has a typical initial transient period, which is resulted from adding the transient process T_t^{LTM}, but due to NE_t, the process will not become ergodic afterward.

We investigated the performance of the Schruben test on the six artificially generated datasets. Table 12 shows that in the case of the linear transient mean process, the Schruben test detected a length close to the pre-defined length of the initial transient period while l was chosen to be less than 300. To explain this further, for example in the case of the linear transient mean process, by changing l from 100 to 1000, the distance between T_t^{LTM} and T_{t+1}^{LTM} reduces from 0.1 to 0.01 and so the bias introduced cannot be picked up by the test. This was easily fixed by increasing the values of x from 10 to 100. In the cases of the linear transient variance processes, the length of the initial transient period detected by the Schruben test was close to zero. The explanation for this is that although there is a transient behaviour of the variance for this process, the mean is constant, and the length of the initial transient period equals zero with a constant mean. The test failed to detect the initialisation bias for the ARMA(5,5) process. As was mentioned before, both periodic and non-ergodic processes do not have a steady state distribution. In our experiment, the Schruben test did not detect a truncation point in the case of the periodic process but failed to detect the non-ergodicity.

Tested Dataset (Number of Replications) \ Length of Initial Transient Period *l*	**50**	**100**	**150**	**200**	**250**	**300**	**1000**
Linear Transient Mean	45.31 (207)	90.65 (122)	137.28 (254)	183.42 (248)	229.45 (1103)	274.08 (2817)	5.69 (553)
Linear Transient Variance	17.22 (862)	7.2 (774)	3.48 (258)	2.01 (164)	1.93 (160)	1.77 (141)	1.09 (64)
Exponential Transient Mean	95.85 (1095)	152.3 (2631)	181.97 (4122)	195.25 (6262)	195.56 (8695)	186.22 (11358)	11.3 (1339)
ARMA(5,5)	0 (50)	0 (50)	0 (50)	0 (50)	0 (50)	0 (50)	0 (50)
Periodic	0 (50)	0 (50)	0 (50)	0 (50)	0 (50)	0 (50)	0 (50)
Non-Ergodic	43.92 (130)	86.72 (125)	127.94 (297)	167.02 (538)	194.99 (2970)	171.38 (10868)	8.82 (681)

Table 12. Performance of the Schruben Test on Six Artificially Generated Processes with Different Lengths of Initial Transient Period.

We also investigated the performance of one of the GSS tests (the Area test) on the artificially generated processes, as all the three GSS tests had performed very similarly in all the previous experiments. As can be seen from Table 13[13], apart from the linear transient variance process, the Area test under-estimated the length of the initial transient period by a large margin. However, in case of the non-ergodic process, the Area test detected the non-ergodicity and did not detect an initialisation point.

[13] As the result of poor detection of the length of the initial transient by the Area test on the artificially generated datasets, we only included *l* = 100 and 200 for this research.

Tested Dataset (Number of Replications) \ Length of Initial Transient Period *l*	100	200
Linear Transient Mean	6.58 (1381)	6.5 (1307)
Linear Transient Variance	93.38 (3524)	127.42 (8260)
Exponential Transient Mean	7.33 (1390)	6.86 (1359)
ARMA(5,5)	5.72 (996)	5.72 (996)
Periodic	7.12 (1432)	7.12 (1432)
Non-Ergodic	0 (50)	0 (50)

Table 13. Performance of the Area Test on Six Artificially Generated processes with Different Lengths of Initial Transient Period.

Lastly, we investigated the performance of the LWS test on the same six artificially generated processes. As Table 14 shows, in the cases of the linear and exponential transient mean and the linear transient variance processes, the LWS test detected a length of initial transient period well over the supposedly length. In case of the non-ergodic process, the LWS test performed well not detecting an initialisation point.

Tested Dataset (Number of Replications) \ Length of Initial Transient Period *l*	**100**	**200**
Linear Transient Mean	1045 (18975)	1877 (23551)
Linear Transient Variance	255.41 (63243)	521.63 (84325)
Exponential Transient Mean	1141 (17423)	1790 (24350)
ARMA(5,5)	0 (50)	0 (50)
Periodic	111.85 (16503)	154.47 (14556)
Non-Ergodic	0 (50)	0 (50)

Table 14. Performance of the LWS Test on Six Artificially Generated processes with Different Lengths of Initial Transient Period.

Therefore, by looking at the results of the performance of these three tests on the artificially generated processes, we can conclude that the Schruben test performs significantly well on the non-queuing systems as well as the three queueing systems examined.

3.2.8. Initialising the System with a Different Number of Customers at Time Zero

Comparing the length of the initial transient period detected by the statistical tests while starting the system with different number of customers at time zero was another issue investigated. Kelton and Law [1994] analysed the transient behaviour of the M/M/1/8 queueing system with an arbitrary number of customers present at time zero. They considered that initialising the simulation in a way that would promote rapid convergence to steady state should shorten the length of the initial transient period.

Following the procedure described in Section 2.1.1, they estimated the length of the initial transient period with $k = 0, \ldots, 35$ number of customers present in the M/M/1/8 queueing system at time zero for $r = 0.5$, 0.8, 0.9 and 0.95 system loads. They showed that with different system loads, the length of the initial transient period falls from $k = 0$ to an apparent minimum then rises with k. They also showed that this minimum value is different for each system load, and for a high r, optimal initialisation provides an impressive shortening of the required initial transient period, compared with empty and idle initialisation. They found that the value of k to reach this minimum length of the initial transient period is slightly larger than L, the mean steady state number of customers in the system. However, the exact value of k to reach this optimal length of initial transient is rarely available in practice, since the mean steady state number of customers in the system (L) can only be calculated for analytically tractable queueing systems. Therefore, they concluded that starting the system in an empty and idle state is more sensible.

To study the influence of the initial state on the length of the initial transient period, we conducted experiments, using the tests in simulations with different numbers of k customers present at time zero. In addition, we chose k relative to the mean steady state expected number of customers; i.e., 0, L, $2L$, and $3L$, where

L is the mean steady state number of customers[14]. The steady state expected number of customers and the steady state expected waiting time (exclusive the service time) in the queue for a customer in the M/M/1/8 system is calculated as $L = l\left(\frac{1}{m} + w\right)$ and $w = \frac{1}{m} + \frac{1}{l}\left(\frac{r^2}{1-r}\right)\left(\frac{1+C_S^2}{2}\right)$, respectively.

Table 15 shows the results of these experiments. Each value in the table is the average taken from a number of independent runs, using the sequential method described in Section 3.1. Contrary to the theoretical results found by Kelton and Law [1994], the length of the initial transient period does not fall to an apparent minimum in our investigation.

We also repeated the same experiments, with a fixed number, k, of customers at time zero (k = 0, …, 10) for system load 0.8 as it was presented in Kelton and Law [1994]. Table 16 shows the result of this experiment. It is interesting to also note that these results are contrary to the results found by Kelton and Law [1994] for system load 0.8 where they showed that the optimal length was found when k = 8. Again, our results show that the length of the initial transient period estimated by the statistical tests generally increase by k and do not fall to an apparent minimum.

[14] k, the number of customers in the system at time zero, must be an integer. Choosing a value of k relative to the mean steady state expected number of customers could result in a fraction. In this case, k was rounded down to the closest integer.

Tests / Load	Schruben Test with a variable window size	Schruben Test with a fixed window size	Batch-mean Test	Maximum Test	Area Test	*k*
***K*=0 number of customers presented in the system at time 0**						
0.05	256.00	256.28	262.28	271.64	267.48	0
0.15	261.70	262.06	268.06	279.50	275.34	0
0.25	271.78	272.48	275.48	285.88	280.68	0
0.35	284.20	284.90	283.90	301.58	295.34	0
0.45	287.84	287.84	302.08	305.20	300.00	0
0.55	309.52	309.80	324.32	325.36	318.08	0
0.65	355.10	346.24	364.54	357.69	350.40	0
0.75	385.83	368.15	385.04	379.30	369.58	0
0.85	542.82	481.51	467.98	471.35	461.82	0
0.95	1016.27	647.16	586.37	591.38	583.44	0
***K*=*L* number of customers presented in the system at time 0**						
0.05	256.80	257.12	263.12	269.36	267.28	0
0.15	263.14	263.48	269.48	281.96	275.72	0
0.25	273.04	273.18	277.18	292.78	283.42	0
0.35	284.42	285.18	284.18	299.78	296.66	0
0.45	289.72	289.72	302.92	303.96	300.84	0
0.55	332.25	329.88	351.41	336.08	336.08	1
0.65	354.61	345.38	375.79	362.94	360.01	1
0.75	421.68	402.56	388.55	389.97	383.04	3
0.85	569.35	503.21	492.11	491.22	484.54	5
0.95	1270.24	720.63	650.79	653.34	644.75	19
***K*=2 ´ *L* number of customers presented in the system at time 0**						
0.05	256.80	257.12	263.12	269.36	267.28	0
0.15	263.14	263.48	269.48	281.96	275.72	0
0.25	273.04	273.18	277.18	292.78	283.42	0
0.35	281.38	281.54	287.54	300.02	293.78	1
0.45	295.92	296.90	301.98	308.22	306.14	1
0.55	322.30	322.90	334.46	331.34	330.30	2
0.65	360.23	353.59	379.98	365.35	362.00	3
0.75	446.77	419.07	436.31	423.89	418.85	6
0.85	656.21	550.97	533.93	534.64	526.24	11
0.95	1598.13	824.82	739.54	743.51	733.66	38
***K*=3 ´ *L* number of customers presented in the system at time 0**						
0.05	256.80	257.12	263.12	269.36	267.28	0
0.15	263.14	263.48	269.48	281.96	275.72	0
0.25	273.04	273.18	277.18	292.78	283.42	0
0.35	281.38	281.54	287.54	300.02	293.78	1
0.45	295.11	295.11	304.16	320.48	312.87	2
0.55	333.98	330.56	336.82	340.18	336.04	3
0.65	464.04	446.54	446.44	449.62	443.38	5
0.75	563.94	489.60	492.37	526.18	509.63	9
0.85	769.24	624.14	597.78	599.82	595.49	16
0.95	1818.75	880.34	769.56	775.08	767.50	57

Table 15. Length of the Initial Transient Period for the M/M/1/8 Queue with Different Number of Customers Presented at Time Zero.

Tests / k	Schruben Test as implemented in Akaroa2	Schruben Test with fixed window length	Batch-mean Test	Maximum Test	Area Test	Rank Test
0	464.51	429.65	421.77	425.88	418.61	2674.11
1	473.08	436.32	428.24	434.23	429.02	2688.48
2	474.18	440.07	429.08	434.80	427.83	2711.68
3	482.09	440.62	428.75	434.07	428.53	2696.16
4	480.81	449.53	436.89	440.89	435.16	2693.45
5	497.46	449.48	445.27	449.45	441.72	2735.21
6	502.11	465.43	456.95	462.74	455.59	2755.93
7	514.05	466.40	457.68	462.16	453.70	2752.46
8	524.36	481.06	469.23	473.53	468.20	2716.53
9	537.16	486.41	480.05	484.38	477.88	2787.82
10	533.69	493.12	483.15	487.29	480.58	2795.84

Table 16. Length of the Initial Transient Period Detected by the Statistical Tests with *k* = 0, …, 10 Number of Customers in the System at Time Zero with traffic load 0.8 for an M/M/1/8 Queue.

Chapter 4

Conclusions

"Initialisation bias can be a major source of error in estimating the steady state value of estimated system performance measure" [Schruben, 1981]. The aim of this research has been to find the most efficient and accurate tests to detect the length of the initial transient period. This investigation included a comparative study of seven candidate tests, the Schruben test with variable window size, the Schruben test with a fixed window size, the Area test, the Maximum test, the Batch-mean test, the Rank test, and the LWS test. We also considered two approximations based on theory, namely, the relaxation time, and the estimation based on the algorithm by Kelton and Law to detect the length of the initial transient period. We compared the length of the initial transient period detected by the statistical tests with these two approximations in three queueing systems: M/M1/8, M/Erlang$_4$/1/8 and M/Pareto$_{a=2.1}$/1/8. We also evaluated the performance of the best of these tests on six artificially generated stochastic processes.

To find the most efficient and accurate tests, three major issues have been investigated in this research:

1. Finding a test that detects a length of the initial transient period close to the ones obtained by the relaxation time and the Kelton and Law algorithm. We performed these comparisons for the three different queueing systems (except for the Rank test that were only examined on the M/M/1/8 model as it detected a very variable length of initial transient period). From this investigation, the Rank test detected a length well over the one approximated by theory. Except for the LWS test, the remaining tests detected a length smaller than the one approximated by the theory in

the heavily loaded systems. An initial transient detector based on the LWS test performed well by detecting a length very close to the one approximated by the relaxation time. However, as was mentioned above, the high variability of the length of initial transient period detected by the LWS test is an issue that should be considered while using this test.

2. Finding a test that detects a length of initial transient period, such that the output data process after this period can be considered as being in steady state. The result of these experiments showed that, the Rank test generally detected the length of the initial transient period when the queue was empty or close to empty. So again, it was obvious that the Rank test did not detect an accurate length of the initial transient period. All the remaining tests performed well in the sense that the observations after the initial transient period detected by theses tests behaved closely to steady state. We also investigated the behaviour of the observations after the initial transient period estimated by the relaxation time and the algorithm by Kelton and Law. The results showed that the behaviour of the observations after the initial transient period detected by a fixed value (such as the ones estimated by the relaxation time and the algorithm by Kelton and Law) is very close to steady state behaviour. However, it should be considered that these estimates are only available for a limited number of queueing systems and cannot be used in general cases.

3. Finding a test that does not produce an extremely variable length of initial transient period. The result of this investigation showed that the Rank test detected extremely variable length of initial transient period, much higher than the other tests. Except for the Rank test, the variability of the length detected by the LWS test was also much higher than the remaining tests. For this reason, we did not include these two tests in our further studies of the M/Erlang$_4$/1/8 and M/Pareto$_{a=2.1}$/1/8 queueing systems. The GSS tests and the Schruben test with the fixed window size gave the best results in not detecting a high variable length of initial transient period.

Finding a test that does not produce an extremely variable length of initial transient period. The result of this investigation showed that the Rank test detected extremely variable length of initial transient period, much higher than the other tests. Except for the Rank test, the variability of the length detected by the LWS test was also much higher than the remaining tests. For this reason, we did not include these two tests in our further studies of the M/Erlang$_4$/1/8 and M/Pareto$_{a=2.1}$/1/8 queueing systems. The GSS tests and the Schruben test with the fixed window size gave the best results in not detecting a high variable length of initial transient period.

In addition to the simulation of the queueing models, we tested the performance of the Schruben, the LWS, and the Area tests on six artificially generated stochastic processes. The result of this investigation showed that the Schruben test performed much better than the other tests in detecting the length of the initial transient period in the case of non-queueing models.

In conclusion, the Rank test had the worst performance in all the experiments performed and cannot be recommended. The LWS test performed well in comparison to the remaining tests in detecting a length close to the one approximated by theory, particularly in the heavily loaded system. However, in addition to the highly variability of the length of initial transient period detected by this test, it did not performed well on the non-queueing models. The results of the remaining tests showed that in heavily loaded systems, the length of the initial transient period detected by these tests were smaller than the one approximated by theory. However, factors like the size of the window, number of batches, and the number of observations in each batch, which have a considerable effect on the outcome of the experiments, should be considered.

Appendix

Von Neumann Randomness Test

Assume that the process $\{X_i\}$ is weakly stationary. The von Neumann test statistic for

H_0: the batch means $\bar{x}_1, \ldots, \bar{x}_b$ are uncorrelated, is

$$C_b = \sqrt{\frac{b^2-2}{b-1}} \times \left(1 - \frac{\sum_{j=1}^{b-1} (\bar{x}_j - \bar{x}_{j+1})^2}{2\sum_{j=1}^{b} (\bar{x}_j - \bar{x}_{m,b})^2} \right)$$

Under H_0, $C_b \approx N(0,1)$ for large k (due to the batch means becoming approximately normal) or for large b (by the Central Limit Theorem). If $\{X_i\}$ has a monotone decreasing autocorrelation function (e.g., the delay process for an M/M/1/8 queueing system), one rejects H_0 at level a, if $C_b > z_1 - a/2$ [Alexopoulos, Fishman, and Seila, 1997].

Kolmogorov-Smirnov Test

Using the Kolmogorov-Smirnov Test, one can determine whether two data sets are drawn from the same distribution. Given the hypothesized continuous distribution function F, this test compares F to the empirical distribution function, *F¢*, of the samples. The Kolmogorov-Smirnov test statistic D is the largest absolute deviation between $F(x)$ and *F¢(x)* over the range of the random variable:

$$D = \max_x \{ | F'(x) - F(x) | \}$$

where *F¢(x)* is defined as

$$F'(x) = \frac{number\ of\ \ samples \le x}{N}$$

where N is the number of samples. For testing against a uniform distribution, we must first sort the samples into ascending order $U_1 \le U_2 \le ... \le U_N$, $(0 \le U_i \le 1$ for all i) then compute the following statistics

$$D^{+} = \max_{1 \le i \le N} \left\{ \left| \frac{i}{N} - U_i \right| \right\}$$

$$D^{-} = \max_{1 \le i \le N} \left\{ \left| U_i - \frac{i-1}{N} \right| \right\}$$

Then $D = \max(D^{+}, D^{-})$. To assess D, we use the hypothesis test. H_0 will be rejected at significance level a if

$$(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}) > C_{1-a}$$

where values of C_{1-a} are given by the following table:

1-*a*	0.850	0.900	0.950	0.975	0.990
C*1-*a	1.138	1.224	1.358	1.480	1.628

The difference between this test and χ^2-test is that, χ^2-test goodness-of-fit test can be applied to discrete distributions such as the binomial and the Poisson. The Kolmogorov-Smirnov test is restricted to continuous distributions.

Moving Average Process

Let N_t be the realisation of an independent normal random process with mean 0 and variance 1 where $t = 1, \ldots, \infty$. The first order moving average process denoted MA(1) is defined as

$$Y_t = c + N_t + dN_{t-1},$$

where c and d could be any constant. This time series is called a first order moving average process, which comes from the fact that Y_t is constructed from a weighted sum, similar to average, of the two most recent values of N [Hamilton, 1994].

Autoregressive Process

Let N_t be the realisation of an independent normal random process with mean 0 and variance 1 where $t = 1, \ldots, \infty$. The first order autoregressive process denoted AR(1) is defined as

$$Y_t = c + N_t + dY_{t-1},$$

where c and d could be any constant [Hamilton, 1994].

References

[1] Alexopoulos, C., Fishman, S., and Seila, A. F. 1997. Computational Experience with the Batch Means Method. Proceedings of the 1997 Winter Simulation Conference, pp. 194-201.

[2] Alexopoulos, C., and Seila, A. F. 1996. Implementing the Batch Means Method in Simulation Experiments. Proceedings of the 1996 Winter Simulation Conference, pp. 214-221.

[3] Alexopoulos, C., and Seila, A. F. 2000. Output Analysis for Simulations. In Proceedings of the 2000 Winter Simulation Conference, Winter Simulation Conference Board of Directors, Orlando, FL, pp. 101-108.

[4] Asmussen, S., Glynn, P. W., and Thorisson, H. 1992. Stationarity Detection in the Initial Transient Problem. Transactions on Modelling and Computer Simulation, Vol. 2, No. 2, pp. 130-157.

[5] Bause, F., and Eickhoff, M. 2002. Initial Transient Period Detection Using Parallel Replications. Proceeding 14th European Simulation Symposium, pp. 85-92.

[6] Bause, F., and Eickhoff, M. 2003. Truncation Point Estimation using Multiple Replications in Parallel. Proceedings of the 2003 Winter Simulation Conference. pp. 414-421.

[7] Cash, C. R., Nelson, B. L., Dippold, D. G., Long, J. M., and Pollard, W. P. 1992. Evaluation of Tests for Initial Condition Bias. In Proceedings of the 1992 Winter Simulation Conference, pp. 577-585.

[8] Chance, F. 1993. A Historical Review of the Initial Transient Problem in Discrete Event Simulation Literature. Technical Report, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

[9] Cohen, J. W. 1969. The Single Server Queue. John Wiley & Sons, New York.

[10] Conway, R. W. 1963. Some Tactical Problems in Digital Simulation. Management Science, Vol. 10, pp. 47-61.

[11] Fishman, G. S. 1972. Bias Consideration in Simulation Experiments. Operations Research, Vol. 20, pp. 685-790.

[12] Fishman, G. S. 1973. Concepts and Methods in Discrete Event Digital Simulation. John Wiley & Sons, New York.

[13] Gafarian, A. V., Ancker, C. J., and Morisaku, T. 1978. Evaluation of Commonly Used Rules for Detecting "Steady state" in Computer Simulation. Naval Research Logistics Quarterly, Vol. 78, pp. 511-529.

[14] Glynn, P. W. 1995. Some New Results on the Initial Transient Problem. In Proceedings of the 1995 Winter Simulation Conference, pp. 165-170.

[15] Glynn, P. W. 1995a. Two Approaches to the Initial Transient Problem. Scientific Computing Lecture Notes in Statistics, Vol. 106, pp. 49-57.

[16] Glynn, P. W., and Iglehart, D. L. 1987. A New Initial Bias Deletion Rule. In Proceedings of the 1987 Winter Simulation Conference, pp. 318-319.

[17] Glynn, P. W., and Heidelberger, P. 1992. Analysis of Initial Transient Deletion for Parallel Steady state Simulations. SIAM Journal on Scientific and Statistical Computing, Vol. 13, No. 4, pp. 904-922.

[18] Glynn, P. W., and Heidelberger, P. 1992a. Experiments with Initial Transient Deletion for Parallel Replicated Steady state Simulations. Management Science, Vol. 38, pp. 400-418.

[19] Goldsman, D., Schruben, L. W., and Swain, J. J. 1994. Tests for Transient Means in Simulated Time Series. Naval Research Logistics, Vol. 41, pp. 171-187.

[20] Goldsman, D., and Tokol, G. 2000. Output Analysis Procedures for Computer Simulations. Proceedings of the 2000 Winter Simulation Conference, pp. 39-45.

[21] Gross, D., and Harris, C. M. 1974. Fundamentals of Queueing Theory. John Wiley & Sons, New York.

[22] Hamilton, J. D. 1994. Time Series Analysis. Princeton University Press, Princeton, New Jersey.

[23] Heidelberger, P., and Welch, P. 1983. Simulation Run Length Control in the Presence of an Initial Transient. Operation Research, Vol. 3, No. 1, pp. 1109-1144.

[24] Hoover, S. V., and Perry, R. F. 1990. Simulation: A Problem-Solving Approach. Addison-Wesley, Boston.

[25] Jackway, P. T., and deSilva, B. M. 1992. A Methodology for Initialisation Bias Reduction in Computer Simulation Output. Asia-Pacific Journal of Operational Research. Vol. 9, pp. 87-100.

[26] Karlin, S., and McGregor, J. L. 1957. The Differential Equations of Birth-and-Death Processes and the Stieltjes Moment Problem. Transactions of the American Mathematical Society. Vol. 85, pp. 489-546.

[27] Kelton, W. D., and Law, A. M. 1983. A New Approach for Dealing with the Start-up Problem in Discrete Event Simulation. Naval Research Logistics Quarterly, Vol. 30, pp. 641-658.

[28] Kelton, W. D., Sadowski, R. P., and Sadowski, D.A. 2002. Simulation with Arena. McGraw-Hill, Boston.

[29] Lada, E. K., Wilson, J. R., and Steiger, N. M. 2003. A Wavelet-based Spectral Method for Steady state Simulation Analysis. Proceedings of the 2003 Winter Simulation Conference, pp. 422-430.

[30] Law, A. M. 1983. Statistical Analysis of Simulation Output Data. Operations Research, Vol. 31, pp. 983-1029.

[31] Law, A. M. and Kelton, W. D. 1991. Simulation Modelling and Analysis, 2nd Edition. New York: McGraw-Hill.

[32] Law, A. M. and Kelton, W. D. 2000. Simulation Modelling and Analysis, 3rd Edition. New York: McGraw-Hill.

[33] Linton, J. R., and Hamonosky. C. M. 2002. A Comparison of Selective Initialisation Bias Elimination Methods. Proceedings of the 2002 Winter Simulation Conference, Vol. 2, pp. 1951-1957.

[34] Ma, X., and Kochhar, A. K. 1993. A Comparison Study of Two Tests for Detecting Initialisation bias in Simulation Output. Simulation, Vol. 61, No. 2, pp. 94-101.

[35] MacDougall, M. H. 1987. Simulating Computer Systems: Techniques and Tools. MIT Press. London.

[36] McNickle, D., Stacey, C., and Pawlikowski, K. 1993. Detection and Significance of the Initial Transient Period in Quantitative Steady state Simulation. Proceedings of the Eighth Australian Teletraffic Research Seminar, pp.193-202.

[37] Molle, J. W., and Morrice, D. J. 1994. Initial Transient Deletion in Simulations Using the Second-order Cumulant Spectrum. Annals of Operations Research, Vol. 53, pp. 443-470.

[38] Morisaku, T. 1976. Techniques for Data-Truncation in Digital Computer Simuation. Ph.D. dissertation. Department of Industrial and Systems Engineering, Universityof Southern California, Los Angeles, California.

[39] Morse, P. M. 1955. Stochastic Properties of Waiting Lines. Operations Research. Vol. 3, pp. 255-261.

[40] Mori, M. 1976. Transient Behaviour of the Mean Waiting Time and Its Exact Forms in M/M/1 and M/D/1. Operations Research Society of Japan. Vol. 19, pp. 14-31.

[41] Newell, G. F.1971. Applications of Queueing Theory. Chapman & Hall, London.

[42] Ockerman, D. H., and Goldsman, D. 1997. The Impact of Transients on Simulation Variance Estimators. Proceedings of the 1997 Winter Simulation Conference, pp. 234-239.

[43] Odoni, A. and Roth, E. 1983. An Empirical Investigation of the Transient Behaviour of Stationary Queueing Systems. Operations Research, Vol. 31, pp. 432-455.

[44] Pawlikowski, K. 1990. Steady state Simulation of Queueing Processes: A Survey of Problems and Solutions. ACM Computing Surveys, Vol. 22, No. 2, pp. 123-170.

[45] Robinson, S. 2002. A Statistical Process Control Approach for Estimating the Warm-up Period. Proceeding of the 2002 Winter Simulation Conference, pp. 439-446.

[46] Schruben, L. 1981. Detecting Initialisation Bias in Simulation Output. Operations Research, Vol. 30, No. 3, pp. 569-590.

[47] Schruben, L. 1982. Detecting Initialisation Bias in Simulation Output. Operations Research, Vol. 30, No. 3, pp. 569-590.

[48] Schruben, L., Singh, H., and Tierney, L. 1983. Optimal Tests for Initialisation Bias in Simulation Output. Operations Research, Vol. 31, No. 6, pp. 1167-1178.

[49] Shannon, R. E. 1975. System Simulation: The Art and Science. Prentice-Hall, Englewood Cliffs, N.J.

[50] Stacey, C. 1993. Application of Independent Replications in Automatic Steady state Simulation. Honours Report, Dept. of Computer Science, University of Canterbury, Christchurch, New Zealand.

[51] Vassilacopoulos, G. 1989. Testing for Initialisation Bias in Simulation Output. Simulation, Vol. 52, pp. 151-153.

[52] Von Neumann, J. 1941. Distribution of the Ratio of the Mean Square Successive Difference to the Variance. Annals of Mathematical Statistics. Vol. 12, pp. 367-395.

[53] White, K. P., Jr. 1995. A Simple Rule for Mitigating Initialisation Bias in Simulation Output: Comparative Results. Proceedings of the 1995 IEEE International Conference on Systems, Man, and Cybernetics. San Antonio. TX, pp. 206-211.

[54] White, K. P., Jr. 1997. An Effective Truncation Heuristic for Bias Reduction in Simulation Output. Simulation, Vol. 69, No. 6, pp. 323-334.

[55] White, K. P., Cobb, M. J., and Spratt, S. C. 2000. A Comparison of Five Steady state Truncation Heuristics for Simulation. Proceedings of the 2000 Winter Simulation Conference, pp. 755-760.

[56] White, K. P., Jr. and Minnox, M. A. 1994. A Simple Rule for Mitigating Initialisation Bias in Simulation Output. Proceedings of the 1994 IEEE International Conference on Systems, Man, and Cybernetics. San Antonio. TX, pp. 219-221.

[57] Welch, P. D. 1983. The Statistical Analysis of Simulation Results. In Computer Performance Modelling Hand Book, Lavenberg Edition, pp. 267-329.

[58] Wilson, J. R. 1984. Statistical Aspects of Simulation. Operations Research, pp. 921-937.

[59] Wilson, J. R., and Pritsker, A. B. 1978. A Survey of Evaluation of Start-up Policies in Simulation Experiments. Simulation, Vol. 31, pp. 79-89.

[60] Yücesan, E. 1993. Randomisation Tests for Initialisation Bias in Simulation Output. Naval Research Logistics, Vol. 40, pp. 643-663.